

Psychological Bulletin

CONTENTS

ARTICLES:

- Construct Validity in Psychological Tests.....LEE J. CRONBACH
AND PAUL E. MEEHL 281
- Psychology in the Arab Near East.....E. TERRY PROTHRO
AND LEVON H. MELIKIAN 303
- The Role of Motivation in Verbal Learning and
Performance.....I. E. FARBER 311
- Learning During Sleep?.....CHARLES W. SIMON
AND WILLIAM H. EMMONS 328
- Teaching in the Ivory Tower, with Rarely a
Step Outside.....S. L. PRESSEY 343
- Additions to Psychological Necrology: Japan, 1928-1952. KOJI SATO 345

SPECIAL REVIEW:

- Handbook of Social Psychology.....DANIEL KATZ 346

BOOK REVIEWS:

- Hoch and Zubin's Depression.....JAMES D. PAGE 354
- Savage's The Foundations of Statistics.....LEONARD S. KOGAN 354
- Roback's Destiny and Motivation in Language.....CALVIN S. HALL 355
- Cooper and Erickson's Time Distortion in
Hypnosis.....ANDRÉ M. WEITZENHOFER 356
- McKeon's Interrelations of Cultures.....CLYDE KLUCKHOHN 357
- Ziskind's Psychophysiologic Medicine.....JOSEPH E. BARMACK 357
- Murphy and Bachrach's An Outline of Abnormal
Psychology.....FRED MCKINNEY 358
- Holjer's Language in Culture.....ROGER W. BROWN 359
- Klopfer, Ainsworth, Klopfer, and Holt's Developments
in the Rorschach Technique. Vol. I. Technique
and Theory.....GEORGE W. ALBEE 360
- Katz, Cartwright, Eldersveld, and Lee's Public
Opinion and Propaganda.....IVAN D. LONDON 361
- McMillan, Grant, Fitts, Frick, McCullough, Miller,
and Brosin's Current Trends in Information Theory.....A. CHAPANIS 361

(Continued on back cover)

Published Bimonthly by the
American Psychological Association

WAYNE DENNIS, Editor

*Brooklyn College, on leave of absence,
The American University, Beirut, Lebanon*

EDWARD GEDEN, Associate Editor (Reviews)
Brooklyn College

ROBERT L. THORNDIKE, Associate Editor (Statistics)
Teachers College, Columbia University

LORRAINE KATZHELET, Managing Editor

Consulting Editors

LAUNCE F. CARTER
*RAND Corporation
Santa Monica, California*

BENTLEY G. GLASS
*Northwestern University
E. Evanston, Illinois*

VICTOR C. RAIMY
University of Colorado

The *Psychological Bulletin* contains evaluative reviews of research reports and essays on research methodology in psychology. This section is devoted to the publication of original research or original theoretical articles.

Manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Manuscripts should be sent to this address during the editor's absence on leave. Manuscripts will be acknowledged at the time of receipt and the carbon copies forwarded to the editor. However, beginning on August 1, 1955 and until August 1, 1956 correspondence should be sent to the editor at the Department of Psychology, American University, Beirut, Lebanon.

Letters concerning book reviews. After this year, book reviews will appear in the JOURNAL but in *Contemporary Psychology*, a new JOURNAL beginning in January, 1956. Hereafter, books for review should be sent to the Editor, Massachusetts Hall, Harvard University, Cambridge 38, Mass.

Preparation of articles for publication. Authors are strongly urged to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 [No. 4, Part 2]). Particular attention should be given to the section on the preparation of manuscripts (pp. 442-443), since this is a particular source of difficulty to the editorial reviewers. All copy must be double spaced, including the references. Manuscripts must be submitted in duplicate. Original figures and photographs must be submitted as photographic or pencil-drawn copies. Authors are urged to retain a copy of the manuscript to guard against loss in the mail. Authors are given 50 free offprints. Fifty free offprints are given to contributors of book reviews. Five copies of the JOURNAL are supplied to authors of book reviews.

Communications—including subscriptions, orders of back volumes, and change of address—should be addressed to the American Psychological Association, 1334 Massachusetts Street N.W., Washington 4, D. C. Address changes must be received at the Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced. Authors are urged to inform their post office that they will guarantee second-class forwarding of the JOURNAL. Undelivered copies must be paid within four months of the date of mailing. Annual subscription: \$8.00 (Foreign \$8.50). Single copies: 50¢.

PRINTED SIMULTANEOUSLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION AND THE AMERICAN PSYCHOLOGICAL SOCIETY

*Washington, Wisconsin
and 1334 Massachusetts Street N.W., Washington 4, D. C.*

Second-class postage paid at the post office at Washington, D. C., and at additional mailing offices. Postmaster: Send address changes in this journal to The American Psychological Association, 1334 Massachusetts Street N.W., Washington 4, D. C. Second-class postage paid at the post office at Washington, D. C., and at additional mailing offices. Postmaster: Send address changes in this journal to The American Psychological Association, 1334 Massachusetts Street N.W., Washington 4, D. C.

Copyright, 1955, by The American Psychological Association

Psychological Bulletin

CONSTRUCT VALIDITY IN PSYCHOLOGICAL TESTS

LEE J. CRONBACH

University of Illinois

AND

PAUL E. MEEHL¹*University of Minnesota*

Validation of psychological tests has not yet been adequately conceptualized, as the APA Committee on Psychological Tests learned when it undertook (1950-54) to specify what qualities should be investigated before a test is published. In order to make coherent recommendations the Committee found it necessary to distinguish four types of validity, established by different types of research and requiring different interpretation. The chief innovation in the Committee's report was the term *construct validity*.² This idea was first formulated by a subcommittee (Meehl and R. C. Challman) studying how proposed recommendations would apply to projective techniques, and later modified and clarified by the entire Committee (Bordin, Challman, Conrad, Humphreys, Super, and the present writers). The statements agreed upon by the Committee (and by committees of two other associations) were published in the *Technical Recommendations* (59). The present interpretation of construct validity is not "official" and deals

with some areas where the Committee would probably not be unanimous. The present writers are solely responsible for this attempt to explain the concept and elaborate its implications.

Identification of construct validity was not an isolated development. Writers on validity during the preceding decade had shown a great deal of dissatisfaction with conventional notions of validity, and introduced new terms and ideas, but the resulting aggregation of types of validity seems only to have stirred the muddy waters. Portions of the distinctions we shall discuss are implicit in Jenkins' paper, "Validity for what?" (33), Gulliksen's "Intrinsic validity" (27), Goodenough's distinction between tests as "signs" and "samples" (22), Cronbach's separation of "logical" and "empirical" validity (11), Guilford's "factorial validity" (25), and Mosier's papers on "face validity" and "validity generalization" (49, 50). Helen Peak (52) comes close to an explicit statement of construct validity as we shall present it.

FOUR TYPES OF VALIDATION

The categories into which the *Recommendations* divide validity studies are: predictive validity, concurrent validity, content validity, and construct validity. The first two of these may be considered together as *criterion-oriented* validation procedures.

The pattern of a criterion-oriented

¹ The second author worked on this problem in connection with his appointment to the Minnesota Center for Philosophy of Science. We are indebted to the other members of the Center (Herbert Feigl, Michael Scriven, Wilfrid Sellars), and to D. L. Thistlethwaite of the University of Illinois, for their major contributions to our thinking and their suggestions for improving this paper.

² Referred to in a preliminary report (58) as *congruent validity*.

study is familiar. The investigator is primarily interested in some criterion which he wishes to predict. He administers the test, obtains an independent criterion measure on the same subjects, and computes a correlation. If the criterion is obtained some time after the test is given, he is studying *predictive validity*. If the test score and criterion score are determined at essentially the same time, he is studying *concurrent validity*. Concurrent validity is studied when one test is proposed as a substitute for another (for example, when a multiple-choice form of spelling test is substituted for taking dictation), or a test is shown to correlate with some contemporary criterion (e.g., psychiatric diagnosis).

Content validity is established by showing that the test items are a sample of a universe in which the investigator is interested. Content validity is ordinarily to be established deductively, by defining a universe of items and sampling systematically within this universe to establish the test.

Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined." The problem faced by the investigator is, "What constructs account for variance in test performance?" Construct validity calls for no new scientific approach. Much current research on tests of personality (9) is construct validation, usually without the benefit of a clear formulation of this process.

Construct validity is not to be identified solely by particular investigative procedures, but by the orientation of the investigator. Criterion-oriented validity, as Bechtoldt emphasizes (3, p. 1245), "involves the *acceptance* of a set of operations as an adequate definition of whatever is to

be measured." When an investigator believes that no criterion available to him is fully valid, he perforce becomes interested in construct validity because this is the only way to avoid the "infinite frustration" of relating every criterion to some more ultimate standard (21). In content validation, *acceptance* of the universe of content as defining the variable to be measured is essential. Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured. Determining what psychological constructs account for test performance is desirable for almost any test. Thus, although the MMPI was originally established on the basis of empirical discrimination between patient groups and so-called normals (concurrent validity), continuing research has tried to provide a basis for describing the personality associated with each score pattern. Such interpretations permit the clinician to predict performance with respect to criteria which have not yet been employed in empirical validation studies (cf. 46, pp. 49-50, 110-111).

We can distinguish among the four types of validity by noting that each involves a different emphasis on the criterion. In predictive or concurrent validity, the criterion behavior is of concern to the tester, and he may have no concern whatsoever with the type of behavior exhibited in the test. (An employer does not care if a worker can manipulate blocks, but the score on the block test may predict something he cares about.) Content validity is studied when the tester is concerned with the type of behavior involved in the test performance. Indeed, if the test is a work sample, the behavior represented in the test may be an end in itself. Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned, and must use indirect measures. Here the trait or quality underlying the test is of central importance, rather than either the test behavior or the scores on the criteria (59, p. 14).

Construct validation is important at times for every sort of psychological test: aptitude, achievement, interests, and so on. Thurstone's statement is interesting in this connection:

In the field of intelligence tests, it used to be common to define validity as the correlation between a test score and some outside criterion. We have reached a stage of sophistication where the test-criterion correlation is too coarse. It is obsolete. If we attempted to ascertain the validity of a test for the second space-factor, for example, we would have to get judges [to] make reliable judgments about people as to this factor. Ordinarily their [the available judges'] ratings would be of no value as a criterion. Consequently, validity studies in the cognitive functions now depend on criteria of internal consistency . . . (60, p. 3).

Construct validity would be involved in answering such questions as: To what extent is this test of intelligence culture-free? Does this test of "interpretation of data" measure reading ability, quantitative reasoning, or response sets? How does a person with A in Strong Accountant, and B in Strong CPA, differ from a person who has these scores reversed?

Example of construct validation procedure. Suppose measure *X* correlates .50 with *Y*, the amount of palmar sweating induced when we tell a student that he has failed a Psychology I exam. Predictive validity of *X* for *Y* is adequately described by the coefficient, and a statement of the experimental and sampling conditions. If someone were to ask, "Isn't there perhaps another way to interpret this correlation?" or "What other kinds of evidence can you bring to support your interpretation?", we would hardly understand what he was asking because no interpretation has been made. These questions become relevant when the correlation is advanced as evidence that "test *X* measures anxiety proneness." Alternative interpretations are possible;

e.g., perhaps the test measures "academic aspiration," in which case we will expect different results if we induce palmar sweating by economic threat. It is then reasonable to inquire about other *kinds* of evidence.

Add these facts from further studies: Test *X* correlates .45 with fraternity brothers' ratings on "tense-ness." Test *X* correlates .55 with amount of intellectual inefficiency induced by painful electric shock, and .68 with the Taylor Anxiety scale. Mean *X* score decreases among four diagnosed groups in this order: anxiety state, reactive depression, "normal," and psychopathic personality. And palmar sweat under threat of failure in Psychology I correlates .60 with threat of failure in mathematics. Negative results eliminate competing explanations of the *X* score; thus, findings of negligible correlations between *X* and social class, vocational aim, and value-orientation make it fairly safe to reject the suggestion that *X* measures "academic aspiration." We can have substantial confidence that *X* does measure anxiety proneness if the current theory of anxiety can embrace the variates which yield positive correlations, and does not predict correlations where we found none.

KINDS OF CONSTRUCTS

At this point we should indicate summarily what we mean by a construct, recognizing that much of the remainder of the paper deals with this question. A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct. We expect a person at any time to possess or not possess a qualitative attribute (amnesia) or structure, or to possess some degree of a quantitative attrib-

ute (cheerfulness). A construct has certain associated meanings carried in statements of this general character: Persons who possess this attribute will, in situation *X*, act in manner *Y* (with a stated probability). The logic of construct validation is invoked whether the construct is highly systematized or loose, used in ramified theory or a few simple propositions, used in absolute propositions or probability statements. We seek to specify how one is to defend a proposed interpretation of a test; *we are not recommending any one type of interpretation.*

The constructs in which tests are to be interpreted are certainly not likely to be physiological. Most often they will be traits such as "latent hostility" or "variable in mood," or descriptions in terms of an educational objective, as "ability to plan experiments." For the benefit of readers who may have been influenced by certain eisegeses of MacCorquodale and Meehl (40), let us here emphasize: Whether or not an interpretation of a test's properties or relations involves questions of construct validity is to be decided by examining the entire body of evidence offered, together with what is asserted about the test in the context of this evidence. Proposed identifications of constructs allegedly measured by the test with constructs of other sciences (e.g., genetics, neuroanatomy, biochemistry) make up only *one* class of construct-validity claims, and a rather minor one at present. Space does not permit full analysis of the relation of the present paper to the MacCorquodale-Meehl distinction between hypothetical constructs and intervening variables. The philosophy of science pertinent to the present paper is set forth later in the section entitled, "The nomological network."

THE RELATION OF CONSTRUCTS TO "CRITERIA"

Critical View of the Criterion Implied

An unquestionable criterion may be found in a practical operation, or may be established as a consequence of an operational definition. Typically, however, the psychologist is unwilling to use the directly operational approach because he is interested in building theory about a generalized construct. A theorist trying to relate behavior to "hunger" almost certainly invests that term with meanings other than the operation "elapsed-time-since-feeding." If he is concerned with hunger as a tissue need, he will not accept time lapse as *equivalent* to his construct because it fails to consider, among other things, energy expenditure of the animal.

In some situations the criterion is no more valid than the test. Suppose, for example, that we want to know if counting the dots on Bender-Gestalt figure five indicates "compulsive rigidity," and take psychiatric ratings on this trait as a criterion. Even a conventional report on the resulting correlation will say something about the extent and intensity of the psychiatrist's contacts and should describe his qualifications (e.g., diplomate status? analyzed?).

Why report these facts? Because data are needed to indicate whether the criterion is any good. "Compulsive rigidity" is not really intended to mean "social stimulus value to psychiatrists." The implied trait involves a range of behavior-dispositions which may be very imperfectly sampled by the psychiatrist. Suppose dot-counting does not occur in a particular patient and yet we find that the psychiatrist has rated him as "rigid." When questioned the psychiatrist tells us that the patient was a rather easy, free-wheeling sort;

however, the patient *did* lean over to straighten out a skewed desk blotter, and this, viewed against certain other facts, tipped the scale in favor of a "rigid" rating. On the face of it, counting Bender dots may be just as good (or poor) a sample of the compulsive-rigidity domain as straightening desk blotters is.

Suppose, to extend our example, we have four tests on the "predictor" side, over against the psychiatrist's "criterion," and find generally positive correlations among the five variables. Surely it is artificial and arbitrary to impose the "test-should-predict-criterion" pattern on such data. The psychiatrist samples verbal content, expressive pattern, voice, posture, etc. The psychologist samples verbal content, perception, expressive pattern, etc. Our proper conclusion is that, from this evidence, the four tests and the psychiatrist all assess some common factor.

The asymmetry between the "test" and the so-designated "criterion" arises only because the terminology of predictive validity has become a commonplace in test analysis. In this study where a construct is the central concern, any distinction between the merit of the test and criterion variables would be justified only if it had already been shown that the psychiatrist's theory and operations were excellent measures of the attribute.

INADEQUACY OF VALIDATION IN TERMS OF SPECIFIC CRITERIA

The proposal to validate constructual interpretations of tests runs counter to suggestions of some others. Spiker and McCandless (57) favor an operational approach. Validation is replaced by compiling statements as to how strongly the test predicts other observed variables of interest. To avoid requiring that each new

variable be investigated completely by itself, they allow two variables to collapse into one whenever the properties of the operationally defined measures are the same: "If a new test is demonstrate¹ to predict the scores on an older, well-established test, then an evaluation of the predictive power of the older test may be used for the new one." But accurate inferences are possible only if the two tests correlate so highly that there is negligible reliable variance in either test, independent of the other. Where the correspondence is less close, one must either retain all the separate variables operationally defined or embark on construct validation.

The practical user of tests must rely on constructs of some generality to make predictions about new situations. Test X could be used to predict palmar sweating in the face of failure without invoking any construct, but a counselor is more likely to be asked to forecast behavior in diverse or even unique situations for which the correlation of test X is unknown. Significant predictions rely on knowledge accumulated around the generalized construct of anxiety. The *Technical Recommendations* state:

It is ordinarily necessary to evaluate construct validity by integrating evidence from many different sources. The problem of construct validation becomes especially acute in the clinical field since for many of the constructs dealt with it is not a question of finding an imperfect criterion but of finding any criterion at all. The psychologist interested in construct validity for clinical devices is concerned with making an estimate of a hypothetical internal process, factor, system, structure, or state and cannot expect to find a clear unitary behavioral criterion. An attempt to identify any one criterion measure or any composite as *the* criterion aimed at is, however, usually unwarranted (59, p. 14-15).

This appears to conflict with arguments for specific criteria prominent at places in the testing literature.

Thus Anastasi (2) makes many statements of the latter character: "It is only as a measure of a specifically defined criterion that a test can be objectively validated at all . . . To claim that a test measures anything over and above its criterion is pure speculation" (p. 67). Yet elsewhere this article supports construct validation. Tests can be profitably interpreted if we "know the relationships between the tested behavior . . . and other behavior samples, none of these behavior samples necessarily occupying the preeminent position of a criterion" (p. 75). Factor analysis with several partial criteria might be used to study whether a test measures a postulated "general learning ability." If the data demonstrate specificity of ability instead, such specificity is "useful in its own right in advancing our knowledge of behavior; it should not be construed as a weakness of the tests" (p. 75).

We depart from Anastasi at two points. She writes, "The validity of a psychological test should not be confused with an analysis of the factors which determine the behavior under consideration." We, however, regard such analysis as a most important type of validation. Second, she refers to "the will-o'-the-wisp of psychological processes which are distinct from performance" (2, p. 77). While we agree that psychological processes are elusive, we are sympathetic to attempts to formulate and clarify constructs which are evidenced by performance but distinct from it. Surely an inductive inference based on a pattern of correlations cannot be dismissed as "pure speculation."

*Specific Criteria Used Temporarily:
The "Bootstraps" Effect*

Even when a test is constructed on the basis of a specific criterion, it may

ultimately be judged to have greater construct validity than the criterion. We start with a vague concept which we associate with certain observations. We then discover empirically that these observations covary with some other observation which possesses greater reliability or is more intimately correlated with relevant experimental changes than is the original measure, or both. For example, the notion of temperature arises because some objects feel hotter to the touch than others. The expansion of a mercury column does not have face validity as an index of hotness. But it turns out that (a) there is a statistical relation between expansion and sensed temperature; (b) observers employ the mercury method with good interobserver agreement; (c) the regularity of observed relations is increased by using the thermometer (e.g., melting points of samples of the same material vary little on the thermometer; we obtain nearly linear relations between mercury measures and pressure of a gas). Finally, (d) a theoretical structure involving unobservable microevents—the kinetic theory—is worked out which explains the relation of mercury expansion to heat. This whole process of conceptual enrichment begins with what in retrospect we see as an extremely fallible "criterion"—the human temperature sense. That original criterion has now been relegated to a peripheral position. We have lifted ourselves by our bootstraps, but in a legitimate and fruitful way.

Similarly, the Binet scale was first valued because children's scores tended to agree with judgments by schoolteachers. If it had not shown this agreement, it would have been discarded along with reaction time and the other measures of ability previously tried. Teacher judgments once constituted the criterion against

which the individual intelligence test was validated. But if today a child's IQ is 135 and three of his teachers complain about how stupid he is, we do not conclude that the test has failed. Quite to the contrary, if no error in test procedure can be argued, we treat the test score as a valid statement about an important quality, and define our task as that of finding out what other variables—personality, study skills, etc.—modify achievement or distort teacher judgment.

EXPERIMENTATION TO INVESTIGATE CONSTRUCT VALIDITY

Validation Procedures

We can use many methods in construct validation. Attention should particularly be drawn to Macfarlane's survey of these methods as they apply to projective devices (41).

Group differences. If our understanding of a construct leads us to expect two groups to differ on the test, this expectation may be tested directly. Thus Thurstone and Chave validated the Scale for Measuring Attitude Toward the Church by showing score differences between church members and nonchurchgoers. Churchgoing is not the criterion of attitude, for the purpose of the test is to measure something other than the crude sociological fact of church attendance; on the other hand, failure to find a difference would have seriously challenged the test.

Only coarse correspondence between test and group designation is expected. Too great a correspondence between the two would indicate that the test is to some degree invalid, because members of the groups are expected to overlap on the test. Intelligence test items are selected initially on the basis of a correspondence to age, but an item that correlates .95

with age in an elementary school sample would surely be suspect.

Correlation matrices and factor analysis. If two tests are presumed to measure the same construct, a correlation between them is predicted. (An exception is noted where some second attribute has positive loading in the first test and negative loading in the second test; then a low correlation is expected. This is a testable interpretation provided an external measure of either the first or the second variable exists.) If the obtained correlation departs from the expectation, however, there is no way to know whether the fault lies in test A, test B, or the formulation of the construct. A matrix of intercorrelations often points out profitable ways of dividing the construct into more meaningful parts, factor analysis being a useful computational method in such studies.

Guilford (26) has discussed the place of factor analysis in construct validation. His statements may be extracted as follows:

"The personnel psychologist wishes to know 'why his tests are valid.' He can place tests and practical criteria in a matrix and factor it to identify 'real dimensions of human personality.' A factorial description is exact and stable; it is economical in explanation; it leads to the creation of pure tests which can be combined to predict complex behaviors." It is clear that factors here function as constructs. Eysenck, in his "criterion analysis" (18), goes farther than Guilford, and shows that factoring can be used explicitly to test hypotheses about constructs.

Factors may or may not be weighted with surplus meaning. Certainly: when they are regarded as "real 'dimensions'" a great deal of surplus meaning is implied, and the interpreter must shoulder a substan-

tial burden of proof. The alternative view is to regard factors as defining a working reference frame, located in a convenient manner in the "space" defined by all behaviors of a given type. Which set of factors from a given matrix is "most useful" will depend partly on predilections, but in essence the best construct is the one around which we can build the greatest number of inferences, in the most direct fashion.

Studies of internal structure. For many constructs, evidence of homogeneity within the test is relevant in judging validity. If a trait such as *dominance* is hypothesized, and the items inquire about behaviors subsumed under this label, then the hypothesis appears to require that these items be generally intercorrelated. Even low correlations, if consistent, would support the argument that people may be fruitfully described in terms of a generalized tendency to dominate or not dominate. The general quality would have power to predict behavior in a variety of situations represented by the specific items. Item-test correlations and certain reliability formulas describe internal consistency.

It is unwise to list uninterpreted data of this sort under the heading "validity" in test manuals, as some authors have done. High internal consistency may *lower* validity. Only if the underlying theory of the trait being measured calls for high item intercorrelations do the correlations support construct validity. Negative item-test correlations may support construct validity, provided that the items with negative correlations are believed irrelevant to the postulated construct and serve as suppressor variables (31, p. 431-436; 44).

Study of distinctive subgroups of items within a test may set an upper limit to construct validity by showing

that irrelevant elements influence scores. Thus a study of the PMA space tests shows that variance can be partially accounted for by a response set, tendency to mark many figures as similar (12). An internal factor analysis of the PEA Interpretation of Data Test shows that in addition to measuring reasoning skills, the test score is strongly influenced by a tendency to say "probably true" rather than "certainly true," regardless of item content (17). On the other hand, a study of item groupings in the DAT Mechanical Comprehension Test permitted rejection of the hypothesis that knowledge about specific topics such as gears made a substantial contribution to scores (13).

Studies of change over occasions. The stability of test scores ("retest reliability," Cattell's "N-technique") may be relevant to construct validation. Whether a high degree of stability is encouraging or discouraging for the proposed interpretation depends upon the theory defining the construct.

More powerful than the retest after uncontrolled intervening experiences is the retest with experimental intervention. If a transient influence swings test scores over a wide range, there are definite limits on the extent to which a test result can be interpreted as reflecting the typical behavior of the individual. These are examples of experiments which have indicated upper limits to test validity: studies of differences associated with the examiner in projective testing, of change of score under alternative directions ("tell the truth" vs. "make yourself look good to an employer"), and of coachability of mental tests. We may recall Gulliksen's distinction (27): When the coaching is of a sort that improves the pupil's intellectual functioning in

school, the test which is affected by the coaching has validity as a measure of intellectual functioning; if the coaching improves test taking but not school performance, the test which responds to the coaching has poor validity as a measure of this construct.

Sometimes, where differences between individuals are difficult to assess by any means other than the test, the experimenter validates by determining whether the test can detect induced intra-individual differences. One might hypothesize that the Zeigarnik effect is a measure of ego involvement, i.e., that with ego involvement there is more recall of incomplete tasks. To support such an interpretation, the investigator will try to induce ego involvement on some task by appropriate directions and compare subjects' recall with their recall for tasks where there was a contrary induction. Sometimes the intervention is drastic. Porteus finds (53) that brain-operated patients show disruption of performance on his maze, but do not show impaired performance on conventional verbal tests and argues therefrom that his test is a better measure of planfulness.

Studies of process. One of the best ways of determining informally what accounts for variability on a test is the observation of the person's process of performance. If it is supposed, for example, that a test measures mathematical competence, and yet observation of students' errors shows that erroneous reading of the question is common, the implications of a low score are altered. Lucas in this way showed that the Navy Relative Movement Test, an aptitude test, actually involved two different abilities: spatial visualization and mathematical reasoning (39).

Mathematical analysis of scoring procedures may provide important

negative evidence on construct validity. A recent analysis of "empathy" tests is perhaps worth citing (14). "Empathy" has been operationally defined in many studies by the ability of a judge to predict what responses will be given on some questionnaire by a subject he has observed briefly. A mathematical argument has shown, however, that the scores depend on several attributes of the judge which enter into his perception of *any* individual, and that they therefore cannot be interpreted as evidence of his ability to interpret cues offered by particular others, or his intuition.

The Numerical Estimate of Construct Validity

There is an understandable tendency to seek a "construct validity coefficient." A numerical statement of the degree of construct validity would be a statement of the proportion of the test score variance that is attributable to the construct variable. This numerical estimate can sometimes be arrived at by a factor analysis, but since present methods of factor analysis are based on linear relations, more general methods will ultimately be needed to deal with many quantitative problems of construct validation.

Rarely will it be possible to estimate definite "construct saturations," because no factor corresponding closely to the construct will be available. One can only hope to set upper and lower bounds to the "loading." If "creativity" is defined as something independent of knowledge, then a correlation of .40 between a presumed test of creativity and a test of arithmetic knowledge would indicate that at least 16 per cent of the reliable test variance is irrelevant to creativity as defined. Laboratory performance on problems such as Maier's "hatrack" would scarcely be

an ideal measure of creativity, but it would be somewhat relevant. If its correlation with the test is .60, this permits a tentative estimate of 36 per cent as a lower bound. (The estimate is tentative because the test might overlap with the irrelevant portion of the laboratory measure.) The saturation seems to lie between 36 and 84 per cent; a cumulation of studies would provide better limits.

It should be particularly noted that rejecting the null hypothesis does not finish the job of construct validation (35, p. 284). The problem is not to conclude that the test "is valid" for measuring the construct variable. The task is to state as definitely as possible the degree of validity the test is presumed to have.

THE LOGIC OF CONSTRUCT VALIDATION

Construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim. The philosophy of science which we believe does most justice to actual scientific practice will now be briefly and dogmatically set forth. Readers interested in further study of the philosophical underpinning are referred to the works by Braithwaite (6, especially Chapter III), Carnap (7; 8, pp. 56-69), Pap (51), Sellars (55, 56), Feigl (19, 20), Beck (4), Kneale (37, pp. 92-110), Hempel (29; 30, Sec. 7).

The Nomological Net

The fundamental principles are these:

1. Scientifically speaking, to "make clear what something is" means to set forth the laws in which

it occurs. We shall refer to the interlocking system of laws which constitute a theory as a *nomological network*.

2. The laws in a nomological network may relate (a) observable properties or quantities to each other; or (b) theoretical constructs to observables; or (c) different theoretical constructs to one another. These "laws" may be statistical or deterministic.

3. A necessary condition for a construct to be scientifically admissible is that it occur in a nomological net, at least *some* of whose laws involve observables. Admissible constructs may be remote from observation, i.e., a long derivation may intervene between the nomologicals which implicitly define the construct, and the (derived) nomologicals of type a. These latter propositions permit predictions about events. The construct is not "reduced" to the observations, but only combined with other constructs in the net to make predictions about observables.

4. "Learning more about" a theoretical construct is a matter of elaborating the nomological network in which it occurs, or of increasing the definiteness of the components. At least in the early history of a construct the network will be limited, and the construct will as yet have few connections.

5. An enrichment of the net such as adding a construct or a relation to theory is justified if it generates nomologicals that are confirmed by observation or if it reduces the number of nomologicals required to predict the same observations. When observations will not fit into the network as it stands, the scientist has a certain freedom in selecting where to modify the network. That is, there may be alternative constructs or ways of organizing the net which for the time being are equally defensible.

6. We can say that "operations"

which are qualitatively very different "overlap" or "measure the same thing" if their positions in the nomological net tie them to the same construct variable. Our confidence in this identification depends upon the amount of inductive support we have for the regions of the net involved. It is not necessary that a direct observational comparison of the two operations be made—we may be content with an intranetwork proof indicating that the two operations yield estimates of the same network-defined quantity. Thus, physicists are content to speak of the "temperature" of the sun and the "temperature" of a gas at room temperature even though the test operations are nonoverlapping because this identification makes theoretical sense.

With these statements of scientific methodology in mind, we return to the specific problem of construct validity as applied to psychological tests. The preceding guide rules should reassure the "toughminded," who fear that allowing construct validation opens the door to nonconfirmable test claims. *The answer is that unless the network makes contact with observations, and exhibits explicit, public steps of inference, construct validation cannot be claimed.* An admissible psychological construct must be behavior-relevant (59, p. 15). For most tests intended to measure constructs, adequate criteria do not exist. This being the case, many such tests have been left unvalidated, or a finespun network of rationalizations has been offered as if it were validation. Rationalization is not construct validation. One who claims that his test reflects a construct cannot maintain his claim in the face of recurrent negative results because these results show that his construct is too loosely defined to yield verifiable inferences.

A rigorous (though perhaps probabilistic) chain of inference is required to establish a test as a measure of a construct. To validate a claim that a test measures a construct, a nomological net surrounding the concept must exist. When a construct is fairly new, there may be few specifiable associations by which to pin down the concept. As research proceeds, the construct sends out roots in many directions, which attach it to more and more facts or other constructs. Thus the electron has more accepted properties than the neutrino; *numerical ability* has more than *the second space factor*.

"Acceptance," which was critical in criterion-oriented and content validities, has now appeared in construct validity. Unless substantially the same nomological net is accepted by the several users of the construct, public validation is impossible. If A uses *aggressiveness* to mean overt assault on others, and B's usage includes repressed hostile reactions, evidence which convinces B that a test measures *aggressiveness* convinces A that the test does not. Hence, the investigator who proposes to establish a test as a measure of a construct must specify his network or theory sufficiently clearly that others can accept or reject it (cf. 41, p. 406). A consumer of the test who rejects the author's theory cannot accept the author's validation. He must validate the test for himself, if he wishes to show that it represents the construct as *he* defines it.

Two general qualifications are in order with reference to the methodological principles 1-6 set forth at the beginning of this section. Both of them concern the amount of "theory," in any high-level sense of that word, which enters into a construct-defining network of laws or lawlike statements. We do not wish

to convey the impression that one always has a very elaborate theoretical network, rich in hypothetical processes or entities.

Constructs as inductive summaries.

In the early stages of development of a construct or even at more advanced stages when our orientation is thoroughly practical, little or no theory in the usual sense of the word need be involved. In the extreme case the hypothesized laws are formulated entirely in terms of descriptive (observational) dimensions although not all of the relevant observations have actually been made.

The hypothesized network "goes beyond the data" only in the limited sense that it purports to *characterize* the behavior facets which belong to an observable but as yet only partially sampled cluster; hence, it generates predictions about hitherto unsampled regions of the phenotypic space. Even though no unobservables or high-order theoretical constructs are introduced, an element of inductive extrapolation appears in the claim that a cluster including some elements not-yet-observed has been identified. Since, as in any sorting or abstracting task involving a finite set of complex elements, several nonequivalent bases of categorization are available, the investigator may choose a hypothesis which generates erroneous predictions. The failure of a supposed, hitherto untried, member of the cluster to behave in the manner said to be characteristic of the group, or the finding that a nonmember of the postulated cluster does behave in this manner, may modify greatly our tentative construct.

For example, one might build an intelligence test on the basis of his background notions of "intellect," including vocabulary, arithmetic calculation, general information, simi-

larities, two-point threshold, reaction time, and line bisection as subtests. The first four of these correlate, and he extracts a huge first factor. This becomes a second approximation of the intelligence construct, described by its pattern of loadings on the four tests. The other three tests have negligible loading on any common factor. On this evidence the investigator reinterprets intelligence as "manipulation of words." Subsequently it is discovered that test-stupid people are rated as unable to express their ideas, are easily taken in by fallacious arguments, and misread complex directions. These data support the "linguistic" definition of intelligence and the test's claim of validity for that construct. But then a block design test with pantomime instructions is found to be strongly saturated with the first factor. Immediately the purely "linguistic" interpretation of Factor I becomes suspect. This finding, taken together with our initial acceptance of the others as relevant to the background concept of intelligence, forces us to reinterpret the concept once again.

If we simply *list* the tests or traits which have been shown to be saturated with the "factor" or which belong to the cluster, no construct is employed. As soon as we even *summarize the properties* of this group of indicators—we are already making some guesses. Intensional characterization of a domain is hazardous since it selects (abstracts) properties and implies that new tests sharing those properties will behave as do the known tests in the cluster, and that tests not sharing them will not.

The difficulties in merely "characterizing the surface cluster" are strikingly exhibited by the use of certain special and extreme groups for purposes of construct validation. The P_d scale of MMPI was originally de-

rived and cross-validated upon hospitalized patients diagnosed "Psychopathic personality, asocial and amoral type" (42). Further research shows the scale to have a limited degree of predictive and concurrent validity for "delinquency" more broadly defined (5, 28). Several studies show associations between P_4 and very special "criterion" groups which it would be ludicrous to identify as "the criterion" in the traditional sense. If one lists these heterogeneous groups and tries to characterize them intensionally, he faces enormous conceptual difficulties. For example, a recent survey of hunting accidents in Minnesota showed that hunters who had "carelessly" shot someone were significantly elevated on P_4 when compared with other hunters (48). This is in line with one's theoretical expectations; when you ask MMPI "experts" to predict for such a group they invariably predict P_4 or M_4 or both. The finding seems therefore to lend some slight support to the construct validity of the P_4 scale. But of course it would be nonsense to *define* the P_4 component "operationally" in terms of, say, accident proneness. We might try to subsume the original phenotype and the hunting-accident proneness under some broader category, such as "Disposition to violate society's rules, whether legal, moral, or just *sensible*." But now we have ceased to have a neat operational criterion, and are using instead a rather vague and wide-range class. Besides, there is worse to come. We want the class specification to cover a group trend that (nondelinquent) high school students judged by their peer group as least "responsible" score over a full sigma higher on P_4 than those judged most "responsible" (23, p. 75). Most of the behaviors contributing to such sociometric choices fall well within

the range of socially permissible action; the proffered criterion specification is still too restrictive. Again, any clinician familiar with MMPI lore would predict an elevated P_4 on a sample of (nondelinquent) professional actors. Chyatte's confirmation of this prediction (10) tends to support *both*: (a) the theory sketch of "what the P_4 factor is, psychologically"; and (b) the claim of the P_4 scale to construct validity for this hypothetical factor. Let the reader try his hand at writing a brief phenotypic criterion specification that will cover both trigger-happy hunters and Broadway actors! And if he should be ingenious enough to achieve this, does his definition also encompass Hovey's report that high P_4 predicts the judgments "not shy" and "unafraid of mental patients" made upon nurses by their supervisors (32, p. 143)? And then we have Gough's report that *low* P_4 is associated with ratings as "good-natured" (24, p. 40), and Roessell's data showing that high P_4 is predictive of "dropping out of high school" (54). The point is that all seven of these "criterion" dispositions would be readily guessed by any clinician having even superficial familiarity with MMPI interpretation; but to mediate these inferences explicitly requires quite a few hypotheses about dynamics, constituting an admittedly sketchy (but far from vacuous) network defining the genotype *psychopathic deviate*.

Vagueness of present psychological laws. This line of thought leads directly to our second important qualification upon the network schema. The idealized picture is one of a tidy set of postulates which jointly entail the desired theorems; since some of the theorems are coordinated to the observation base, the system constitutes an implicit definition of the

theoretical primitives and gives them an indirect empirical meaning. In practice, of course, even the most advanced physical sciences only approximate this ideal. Questions of "categoricalness" and the like, such as logicians raise about pure calculi, are hardly even statable for empirical networks. (What, for example, would be the desiderata of a "well-formed formula" in molar behavior theory?) Psychology works with crude, half-explicit formulations. We do not worry about such advanced formal questions as "whether all molar-behavior statements are decidable by appeal to the postulates" because we know that no existing theoretical network suffices to predict even the *known* descriptive laws. Nevertheless, the sketch of a network is there; if it were not, we would not be saying *anything* intelligible about our constructs. We do not have the rigorous implicit definitions of formal calculi (which still, be it noted, usually permit of a multiplicity of interpretations). Yet the vague, avowedly incomplete network still gives the constructs whatever meaning they do have. When the network is very incomplete, having many strands missing entirely and some constructs tied in only by tenuous threads, then the "implicit definition" of these constructs is disturbingly loose; one might say that the meaning of the constructs is underdetermined. *Since the meaning of theoretical constructs is set forth by stating the laws in which they occur, our incomplete knowledge of the laws of nature produces a vagueness in our constructs* (see Hempel, 30; Kaplan, 34; Pap, 51). We will be able to say "what anxiety is" when we know all of the laws involving it; meanwhile, since we are in the process of discovering these laws, we do not yet know precisely what anxiety is.

CONCLUSIONS REGARDING THE NETWORK AFTER EXPERIMENTATION

The proposition that x per cent of test variance is accounted for by the construct is inserted into the accepted network. The network then generates a testable prediction about the relation of the test scores to certain other variables, and the investigator gathers data. If prediction and result are in harmony, he can retain his belief that the test measures the construct. The construct is at best adopted, never demonstrated to be "correct."

We do not first "prove" the theory, and then validate the test, nor conversely. In any probable inductive type of inference from a pattern of observations, we examine the relation between the total network of theory and observations. The system involves propositions relating test to construct, construct to other constructs, and finally relating some of these constructs to observables. In ongoing research the chain of inference is very complicated. Kelly and Fiske (36, p. 124) give a complex diagram showing the numerous inferences required in validating a prediction from assessment techniques, where theories about the criterion situation are as integral a part of the prediction as are the test data. A predicted empirical relationship permits us to test all the propositions leading to that prediction. Traditionally the proposition claiming to interpret the test has been set apart as the hypothesis being tested, but actually the evidence is significant for all parts of the chain. If the prediction is not confirmed, any link in the chain may be wrong.

A theoretical network can be divided into subtheories used in making particular predictions. All the events successfully predicted through a subtheory are of course evidence in favor of that theory. Such a subtheory

may be so well confirmed by voluminous and diverse evidence that we can reasonably view a particular experiment as relevant only to the test's validity. If the theory, combined with a proposed test interpretation, mispredicts in this case, it is the latter which must be abandoned. On the other hand, the accumulated evidence for a test's construct validity may be so strong that an instance of misprediction will force us to modify the subtheory employing the construct rather than deny the claim that the test measures the construct.

Most cases in psychology today lie somewhere between these extremes. Thus, suppose we fail to find a greater incidence of "homosexual signs" in the Rorschach records of paranoid patients. Which is more strongly disconfirmed—the Rorschach signs or the orthodox theory of paranoia? The negative finding shows the bridge between the two to be undependable, but this is all we can say. The bridge cannot be used unless one end is placed on solid ground. The investigator must decide which end it is best to relocate.

Numerous successful predictions dealing with phenotypically diverse "criteria" give greater weight to the claim of construct validity than do fewer predictions, or predictions involving very similar behaviors. In arriving at diverse predictions, the hypothesis of test validity is connected each time to a subnetwork largely independent of the portion previously used. Success of these derivations testifies to the inductive power of the test-validity statement, and renders it unlikely that an equally effective alternative can be offered.

Implications of Negative Evidence

The investigator whose prediction and data are discordant must make

strategic decisions. His result can be interpreted in three ways:

1. The test does not measure the construct variable.

2. The theoretical network which generated the hypothesis is incorrect.

3. The experimental design failed to test the hypothesis properly. (Strictly speaking this may be analyzed as a special case of 2, but in practice the distinction is worth making.)

For further research. If a specific fault of procedure makes the third a reasonable possibility, his proper response is to perform an adequate study, meanwhile making no report. When faced with the other two alternatives, he may decide that his test does not measure the construct adequately. Following that decision, he will perhaps prepare and validate a new test. Any rescoring or new interpretative procedure for the original instrument, like a new test, requires validation *by means of a fresh body of data.*

The investigator may regard interpretation 2 as more likely to lead to eventual advances. It is legitimate for the investigator to call the network defining the construct into question, if he has confidence in the test. Should the investigator decide that some step in the network is unsound, he may be able to invent an alternative network. Perhaps he modifies the network by splitting a concept into two or more portions, e.g., by designating types of *anxiety*, or perhaps he specifies added conditions under which a generalization holds. When an investigator modifies the theory in such a manner, he is now required to *gather a fresh body of data* to test the altered hypotheses. This step should normally precede publication of the modified theory. If the new data are consistent with the modified network, he is free from the fear that

his nomologicals were gerrymandered to fit the peculiarities of his first sample of observations. He can now trust his test to some extent, because his test results behave as predicted.

The choice among alternatives, like any strategic decision, is a gamble as to which course of action is the best investment of effort. Is it wise to modify the theory? That depends on how well the system is confirmed by prior data, and how well the modifications fit available observations. Is it worth while to modify the test in the hope that it will fit the construct? That depends on how much evidence there is—apart from this abortive experiment—to support the hope, and also on how much it is worth to the investigator's ego to salvage the test. The choice among alternatives is a matter of research planning.

For practical use of the test. The consumer can accept a test as a measure of a construct only when there is a strong positive fit between predictions and subsequent data. When the evidence from a proper investigation of a published test is essentially negative, it should be reported as a stop sign to discourage use of the test pending a reconciliation of test and construct, or final abandonment of the test. If the test has not been published, it should be restricted to research use until some degree of validity is established (1). The consumer can await the results of the investigator's gamble with confidence that proper application of the scientific method will ultimately tell whether the test has value. Until the evidence is in, he has no justification for employing the test as a basis for terminal decisions. The test may serve, at best, only as a source of suggestions about individuals to be confirmed by other evidence (15, 47).

There are two perspectives in test validation. From the viewpoint of

the psychological practitioner, the burden of proof is on the test. A test should not be used to measure a trait until its proponent establishes that predictions made from such measures are consistent with the best available theory of the trait. In the view of the test developer, however, both the test and the theory are under scrutiny. He is free to say to *himself privately*, "If my test disagrees with the theory, so much the worse for the theory." This way lies delusion, unless he continues his research using a better theory.

Reporting of Positive Results

The test developer who finds positive correspondence between his proposed interpretation and data is expected to report the basis for his validity claim. Defending a claim of construct validity is a major task, not to be satisfied by a discourse without data. The *Technical Recommendations* have little to say on reporting of construct validity. Indeed, the only detailed suggestions under that heading refer to correlations of the test with other measures, together with a cross reference to some other sections of the report. The two key principles, however, call for the most comprehensive type of reporting. The manual for any test "should report all available information which will assist the user in determining what psychological attributes account for variance in test scores" (59, p. 27). And, "The manual for a test which is used primarily to assess postulated attributes of the individual should outline the theory on which the test is based and organize whatever partial validity data there are to show in what way they support the theory" (59, p. 28). It is recognized, by a classification as "very desirable" rather than "essential," that the latter recom-

mendation goes beyond present practice of test authors.

The proper goals in reporting construct validation are to make clear (a) what interpretation is proposed, (b) how adequately the writer believes this interpretation is substantiated, and (c) what evidence and reasoning lead him to this belief. Without *a* the construct validity of the test is of no use to the consumer. Without *b* the consumer must carry the entire burden of evaluating the test research. Without *c* the consumer or reviewer is being asked to take *a* and *b* on faith. The test manual cannot always present an exhaustive statement on these points, but it should summarize and indicate where complete statements may be found.

To specify the interpretation, the writer must state what construct he has in mind, and what meaning he gives to that construct. For a construct which has a short history and has built up few connotations, it will be fairly easy to indicate the presumed properties of the construct, i.e., the nomologicals in which it appears. For a construct with a longer history, a summary of properties and references to previous theoretical discussions may be appropriate. It is especially critical to distinguish proposed interpretations from other meanings previously given the same construct. The validator faces no small task; he must somehow communicate a theory to his reader.

To evaluate his evidence calls for a statement like the conclusions from a program of research, noting what is well substantiated and what alternative interpretations have been considered and rejected. The writer must note what portions of his proposed interpretation are speculations, extrapolations, or conclusions from insufficient data. The author has an

ethical responsibility to prevent unsubstantiated interpretations from appearing as truths. A claim is unsubstantiated unless the evidence for the claim is public, so that other scientists may review the evidence, criticize the conclusions, and offer alternative interpretations.

The report of evidence in a test manual must be as complete as any research report, except where adequate public reports can be cited. Reference to something "observed by the writer in many clinical cases" is worthless as evidence. Full case reports, on the other hand, may be a valuable source of evidence so long as these cases are representative and negative instances receive due attention. The report of evidence must be interpreted with reference to the theoretical network in such a manner that the reader sees why the author regards a particular correlation or experiment as confirming (or throwing doubt upon) the proposed interpretation. Evidence collected by others must be taken fairly into account.

VALIDATION OF A COMPLEX TEST "AS A WHOLE"

Special questions must be considered when we are investigating the validity of a test which is aimed to provide information about several constructs. In one sense, it is naive to inquire "Is this test valid?" One does not validate a test, but only a principle for making inferences. If a test yields many different types of inferences, some of them can be valid and others invalid (cf. Technical Recommendation C2: "The manual should report the validity of each type of inference for which a test is recommended"). From this point of view, every topic sentence in the typical book on Rorschach interpretation presents a hypothesis re-

quiring validation, and one should validate inferences about each aspect of the personality separately and in turn, just as he would want information on the validity (concurrent or predictive) for each scale of MMPI.

There is, however, another defensible point of view. If a test is purely empirical, based strictly on observed connections between response to an item and some criterion, then of course the validity of one scoring key for the test does not make validation for its other scoring keys any less necessary. But a test may be developed on the basis of a theory which in itself provides a linkage between the various keys and the various criteria. Thus, while Strong's Vocational Interest Blank is developed empirically, it also rests on a "theory" that a youth can be expected to be satisfied in an occupation if he has interests common to men now happy in the occupation. When Strong finds that those with high Engineering interest scores in college are preponderantly in engineering careers 19 years later, he has partly validated the proposed use of the Engineer score (predictive validity). Since the evidence is consistent with the theory on which all the test keys were built, this evidence alone increases the presumption that the *other* keys have predictive validity. How strong is this presumption? Not very, from the viewpoint of the traditional skepticism of science. Engineering interests may stabilize early, while interests in art or management or social work are still unstable. A claim cannot be made that the whole Strong approach is valid just because one score shows predictive validity. But if thirty interest scores were investigated longitudinally and all of them showed the type of validity predicted by Strong's theory, we would indeed be caviling to say that this evidence

gives no confidence in the long-range validity of the thirty-first score.

Confidence in a theory is increased as more relevant evidence confirms it, but it is always possible that tomorrow's investigation will render the theory obsolete. The Technical Recommendations suggest a rule of reason, and ask for evidence for each *type* of inference for which a test is recommended. It is stated that no test developer can present predictive validities for all possible criteria; similarly, no developer can run all possible experimental tests of his proposed interpretation. But the recommendation is more subtle than advice that a lot of validation is better than a little.

Consider the Rorschach test. It is used for many inferences, made by means of nomological networks at several levels. At a low level are the simple unrationalized correspondences presumed to exist between certain signs and psychiatric diagnoses. Validating such a sign does nothing to substantiate Rorschach theory. For other Rorschach formulas an explicit *a priori* rationale exists (for instance, high *F%* interpreted as implying rigid control of impulses). Each time such a sign shows correspondence with criteria, its rationale is supported just a little. At a still higher level of abstraction, a considerable body of theory surrounds the general area of *outer control*, interlacing many different constructs. As evidence cumulates, one should be able to decide what specific inference-making chains within this system can be depended upon. One should also be able to conclude—or deny—that so much of the system has stood up under test that one has some confidence in even the untested lines in the network.

In addition to relatively delimited nomological networks surrounding

control or aspiration, the Rorschach interpreter usually has an overriding theory of the test as a whole. This may be a psychoanalytic theory, a theory of perception and set, or a theory stated in terms of learned habit patterns. Whatever the theory of the interpreter, whenever he validates an inference from the system, he obtains some reason for added confidence in his overriding system. His total theory is not tested, however, by experiments dealing with only one limited set of constructs. The test developer must investigate far-separated, independent sections of the network. The more diversified the predictions the system is required to make, the greater confidence we can have that only minor parts of the system will later prove faulty. Here we begin to glimpse a logic to defend the judgment that the test and its whole interpretative system is valid at some level of confidence.

There are enthusiasts who would conclude from the foregoing paragraphs that since there is some evidence of correct, diverse predictions made from the Rorschach, the test as a whole can now be accepted as validated. This conclusion overlooks the negative evidence. Just one finding contrary to expectation, based on sound research, is sufficient to wash a whole theoretical structure away. Perhaps the remains can be salvaged to form a new structure. But this structure now must be exposed to fresh risks, and sound negative evidence will destroy it in turn. There is sufficient negative evidence to prevent acceptance of the Rorschach and its accompanying interpretative structures as a whole. So long as any aspects of the overriding theory stated for the test have been disconfirmed, this structure must be rebuilt.

Talk of areas and structures may seem not to recognize those who

would interpret the personality "globally." They may argue that a test is best validated in matching studies. Without going into detailed questions of matching methodology, we can ask whether such a study validates the nomological network "as a whole." The judge does employ some network in arriving at his conception of his subject, integrating specific inferences from specific data. Matching studies, if successful, demonstrate only that each judge's interpretative theory has some validity, that it is not completely a fantasy. Very high consistency between judges is required to show that they are using the same network, and very high success in matching is required to show that the network is dependable.

If inference is less than perfectly dependable, we must know which aspects of the interpretative network are least dependable and which are most dependable. Thus, even if one has considerable confidence in a test "as a whole" because of frequent successful inferences, one still returns as an ultimate aim to the request of the Technical Recommendation for separate evidence on the validity of each type of inference to be made.

RECAPITULATION

Construct validation was introduced in order to specify types of research required in developing tests for which the conventional views on validation are inappropriate. Personality tests, and some tests of ability, are interpreted in terms of attributes for which there is no adequate criterion. This paper indicates what sorts of evidence can substantiate such an interpretation, and how such evidence is to be interpreted. The following points made in the discussion are particularly significant.

1. A construct is defined implicitly by a network of associations or propo-

sitions in which it occurs. Constructs employed at different stages of research vary in definiteness.

2. Construct validation is possible only when some of the statements in the network lead to predicted relations among observables. While some observables may be regarded as "criteria," the construct validity of the criteria themselves is regarded as under investigation.

3. The network defining the construct, and the derivation leading to the predicted observation, must be reasonably explicit so that validating evidence may be properly interpreted.

4. Many types of evidence are relevant to construct validity, including content validity, interitem correlations, intertest correlations, test-"criterion" correlations, studies of stability over time, and stability under experimental intervention. High correlations and high stability may constitute either favorable or unfavorable evidence for the proposed interpretation, depending on the theory surrounding the construct.

5. When a predicted relation fails to occur, the fault may lie in the proposed interpretation of the test or in the network. Altering the network so that it can cope with the new observations is, in effect, redefining the construct. Any such new interpretation of the test must be validated by a fresh body of data before being advanced publicly. Great care is required to avoid substituting a posteriori rationalizations for proper validation.

6. Construct validity cannot generally be expressed in the form of a single simple coefficient. The data often permit one to establish upper and lower bounds for the proportion of test variance which can be attributed to the construct. The integration of diverse data into a proper interpretation cannot be an entirely quantitative process.

7. Constructs may vary in nature from those very close to "pure description" (involving little more than extrapolation of relations among observation-variables) to highly theoretical constructs involving hypothesized entities and processes, or making identifications with constructs of other sciences.

8. The investigation of a test's construct validity is not essentially different from the general scientific procedures for developing and confirming theories.

Without in the least *advocating* construct validity as preferable to the other three kinds (concurrent, predictive, content), we do believe it imperative that psychologists make a place for it in their methodological thinking, so that its rationale, its scientific legitimacy, and its dangers may become explicit and familiar. This would be preferable to the widespread current tendency to engage in what actually amounts to construct validation research and use of constructs in practical testing, while talking an "operational" methodology which, if adopted, would force research into a mold it does not fit.

REFERENCES

1. AMERICAN PSYCHOLOGICAL ASSOCIATION. *Ethical standards of psychologists*. Washington, D.C.: American Psychological Association, Inc., 1953.
2. ANASTASI, ANNE. The concept of validity in the interpretation of test scores. *Educ. psychol. Measmt*, 1950, 10, 67-78.
3. BECHTOLDT, H. P. Selection. In S. S' Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1237-1267.
4. BECK, L. W. Constructions and inferred entities. *Phil. Sci.*, 1950, 17. Reprinted in H. Feigl and M. Brodbeck

- (Eds.), *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953. Pp. 368-381.
5. BLAIR, W. R. N. A comparative study of disciplinary offenders and non-offenders in the Canadian Army. *Canad. J. Psychol.*, 1950, 4, 49-62.
 6. BRAITHWAITE, R. B. *Scientific explanation*. Cambridge: Cambridge Univer. Press, 1953.
 7. CARNAP, R. Empiricism, semantics, and ontology. *Rév. int. de Phil.*, 1950, II, 20-40. Reprinted in P. P. Wiener (Ed.), *Readings in philosophy of science*, New York: Scribner's, 1953. Pp. 509-521.
 8. CARNAP, R. *Foundations of logic and mathematics*. *International encyclopedia of unified science*, I, No. 3. Pages 56-69 reprinted as "The interpretation of physics" in H. Feigl and M. Brodbeck (Eds.), *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953. Pp. 309-318.
 9. CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, 5, 149-171.
 10. CHVATTE, C. Psychological characteristics of a group of professional actors. *Occupations*, 1949, 27, 245-250.
 11. CRONBACH, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
 12. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
 13. CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297-335.
 14. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, 52, 177-193.
 15. CRONBACH, L. J. The counselor's problems from the perspective of communication theory. In Vivian H. Hower (Ed.), *New perspectives in counseling*. Minneapolis: Univer. of Minnesota Press, 1955.
 16. CURETON, E. E. Validity. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, D. C.: American Council on Education, 1950. Pp. 621-695.
 17. DAMRIN, DORA E. A comparative study of information derived from a diagnostic problem-solving test by logical and factorial methods of scoring. Unpublished doctor's dissertation, Univer. of Illinois, 1952.
 18. EYSENCK, H. J. Criterion analysis—an application of the hypothetico-deductive method in factor analysis. *Psychol. Rev.*, 1950, 57, 38-53.
 19. FEIGL, H. Existential hypotheses. *Phil. Sci.*, 1950, 17, 35-62.
 20. FEIGL, H. Confirmability and confirmation. *Rév. int. de Phil.*, 1951, 5, 1-12. Reprinted in P. P. Wiener (Ed.), *Readings in philosophy of science*. New York: Scribner's, 1953. Pp. 522-530.
 21. GAYLORD, R. H. Conceptual consistency and criterion equivalence: a dual approach to criterion analysis. Unpublished manuscript (PRB Research Note No. 17). Copies obtainable from ASTIA-DSC, AD-21 440.
 22. GOODENOUGH, FLORENCE L. *Mental testing*. New York: Rinehart, 1950.
 23. GOUGH, H. G., MCCLOSKEY, H., & MEEHL, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, 47, 73-80.
 24. GOUGH, H. G., MCKEE, M. G., & YANDELL, R. J. Adjective check list analyses of a number of selected psychometric and assessment variables. Unpublished manuscript. Berkeley: IPAR, 1953.
 25. GUILFORD, J. P. New standards for test evaluation. *Educ. psychol. Measmt.*, 1946, 6, 427-439.
 26. GUILFORD, J. P. Factor analysis in a test-development program. *Psychol. Rev.*, 1948, 55, 79-94.
 27. GULLIKSEN, H. Intrinsic validity. *Amer. Psychologist*, 1950, 5, 511-517.
 28. HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: Univer. of Minnesota Press, 1953.
 29. HEMPEL, C. G. Problems and changes in the empiricist criterion of meaning. *Rév. int. de Phil.*, 1950, 4, 41-63. Reprinted in L. Linsky, *Semantics and the philosophy of language*. Urbana: Univer. of Illinois Press, 1952. Pp. 163-185.
 30. HEMPEL, C. G. *Fundamentals of concept formation in empirical science*. Chicago: Univer. of Chicago Press, 1952.
 31. HORST, P. The prediction of personal adjustment. *Soc. Sci. Res. Council Bull.*, 1941. No. 48.
 32. HOVEY, H. B. MMPI profiles and personality characteristics. *J. consult. Psychol.*, 1953, 17, 142-146.
 33. JENKINS, J. G. Validity for what? *J. consult. Psychol.*, 1946, 10, 93-98.
 34. KAPLAN, A. Definition and specification of meaning. *J. Phil.*, 1946, 43, 281-288.
 35. KELLY, E. L. Theory and techniques of assessment. *Annu. Rev. Psychol.*, 1954, 5, 281-311.
 36. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psy-*

- chology. Ann Arbor: Univer. of Michigan Press, 1951.
37. KNEALE, W. *Probability and induction*. Oxford: Clarendon Press, 1949. Pages 92-110 reprinted as "Induction, explanation, and transcendent hypotheses" in H. Feigl and M. Brodbeck (Eds.), *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953. Pp. 353-367.
 38. LINDQUIST, E. F. *Educational measurement*. Washington, D. C.: American Council on Education, 1950.
 39. LUCAS, C. M. Analysis of the relative movement test by a method of individual interviews. *Bur. Naval Personnel Res. Rep.*, Contract Nonr-694 (00), NR 151-13, Educational Testing Service, March 1953.
 40. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.
 41. MACFARLANE, JEAN W. Problems of validation inherent in projective methods. *Amer. J. Orthopsychiat.*, 1942, 12, 405-410.
 42. MCKINLEY, J. C., & HATHAWAY, S. R. The MMPI: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153-174.
 43. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1948, 12, 20-31.
 44. MEEHL, P. E. A simple algebraic development of Horst's suppressor variables. *Amer. J. Psychol.*, 1945, 58, 550-554.
 45. MEEHL, P. E. An investigation of a general normality or control factor in personality testing. *Psychol. Monogr.*, 1945, 59, No. 4 (Whole No. 274).
 46. MEEHL, P. E. *Clinical vs. statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
 47. MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.*, 1955, 52, 194-216.
 48. Minnesota Hunter Casualty Study. St. Paul: Jacob Schmidt Brewing Company, 1954.
 49. MOSIER, C. I. A critical examination of the concepts of face validity. *Educ. psychol. Measmt*, 1947, 7, 191-205.
 50. MOSIER, C. I. Problems and designs of cross-validation. *Educ. psychol. Measmt*, 1951, 11, 5-12.
 51. PAP, A. Reduction-sentences and open concepts. *Methodos*, 1953, 5, 3-30.
 52. PEAK, HELEN. Problems of objective observation. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden Press, 1953. Pp. 243-300.
 53. PORTEUS, S. D. *The Porteus maze test and intelligence*. Palo Alto: Pacific Books, 1950.
 54. ROESSEL, F. P. MMPI results for high school drop-outs and graduates. Unpublished doctor's dissertation, Univer. of Minnesota, 1954.
 55. SELLARS, W. S. Concepts as involving laws and inconceivable without them. *Phil. Sci.*, 1948, 15, 287-315.
 56. SELLARS, W. S. Some reflections on language games. *Phil. Sci.*, 1954, 21, 204-228.
 57. SPIKER, C. C., & McCANDLESS, B. R. The concept of intelligence and the philosophy of science. *Psychol. Rev.*, 1954, 61, 255-267.
 58. Technical recommendations for psychological tests and diagnostic techniques: preliminary proposal. *Amer. Psychologist*, 1952, 7, 461-476.
 59. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull. Supplement*, 1954, 51, 2, Part 2, 1-38.
 60. THURSTONE, L. L. The criterion problem in personality research. *Psychometric Lab. Rep.*, No. 78. Chicago: Univer. of Chicago, 1952.

Received for early publication February 18, 1955

PSYCHOLOGY IN THE ARAB NEAR EAST

E. TERRY PROTHRO AND LEVON H. MELIKIAN

American University of Beirut

It is the purpose of this paper to present a general picture of psychology in the Arab Near East. Particular attention will be given to places and institutions where most of the psychological activity is concentrated, the type of work carried on in each of those places, and persons who seem to be leaders in the field. The emphasis is on the contemporary scene, but background factors are introduced where they seem to shed light on the current situation.

There is some disagreement among scholars on the inclusiveness of the terms "Near East" and "Middle East." We have included Egypt, Lebanon, Syria, Iraq, Jordan, Saudi Arabia, Bahrein Island, and Kuwait under the phrase "Arab Near East." Some British writers use "Arab Middle East" to refer to the same area. The population of this region is approximately 40 million, of which about half is in Egypt. The total area is more than 1,600,000 square miles. Thus the region is about one fourth as populous, and more than half as large, as the continental United States. Population density is considerably greater than these figures would suggest, for well over three-fourths of the area is desert and the population is concentrated in the fertile, watered areas.

The recorded political history of the Near East goes back thousands of years. We shall not attempt to trace the evolution of these countries to their present form. It is sufficient here to note that all are independent sovereign nations except Bahrein and Kuwait. These two are British protectorates.

The information presented in this paper is based principally on first-hand observation and personal interviews by the authors. The senior author has taught in the Near East for more than two years. The junior author was born in the Near East and has spent most of his life there. Preliminary information about psychology in the area was gathered through correspondence and informal discussions. When it was decided to make an intensive survey of the area, information was solicited by mail from all known psychologists in each country. On the basis of this information, interviews were scheduled in the spring of 1953 with leading psychologists in each country. One or both of us then visited each state, talked at length with the selected psychologists, and personally inspected laboratories, libraries, classrooms, etc. Visits were not made to Kuwait, Bahrein, or Saudi Arabia because we knew of no psychologists there. For information about these countries we relied on the statements of educators from them who visited Lebanon, and on the report of Dr. Habib Kurani, who inspected some of their educational facilities in 1952. Because of the relative importance of Egyptian psychology, additional information about work there was obtained through the kind cooperation of Dr. Hanna Rizk of the American University in Cairo. He prepared and circulated a detailed questionnaire to all Egyptians engaged in full or part-time psychological work. From the returned questionnaires we were able to get an unusually complete picture of psychology in Egypt.

GENERAL CHARACTERISTICS

Psychology in the Arab Near East sprang from and is nourished by Western psychology. British, French, and American influences can be seen—and distinguished—in the work of psychologists throughout the area. The kind and strength of these influences is determined by several factors: the nature of Arab culture, the social and political forces exerted by Western nations, and the comparative vigor of the several transplanted psychologies. Consequently Near Eastern psychology is of interest not only for its own sake, but also as a field for the study of forces affecting the development of various psychological interests and viewpoints.

In the days of the Ottoman Empire, the French school system became the model for Arab schools. With the breakup of that empire in World War I, French influence gained in the Levant but schools of the British type began to flourish in Iraq, Jordan, Palestine, and Egypt; today government schools still follow French and British organizational patterns.

In Syria, Lebanon, and Palestine there are many private schools which were established by Western missionaries before the Ottomans created any local schools. The private schools with foreign support continue to exercise great influence on the education of the area. Consequently the Arab who begins his study of psychology in his own land, whether in public or private colleges, will approach the subject from a Western point of view. If he seeks graduate training, he will probably go to France, Britain, or the United States.

There is remarkably little communication among psychologists of the area. Dependence on Western universities for advanced training

reduces the amount of contact among Arabs of different nationalities. A French-trained psychologist in Lebanon may not read the journal published by a French-trained psychologist in Egypt, although he will be quite *au courant* of the publications from Paris and Lyon. A British-trained psychologist in Damascus may know nothing of the work of the Baghdad psychologist who also received his doctorate in England. This isolation is augmented by severe travel restrictions imposed by many of the Arab states. Since there is little integration of psychological activities in this area, we shall deal with psychology in each of the several countries in turn. The order of presentation is based on our judgment of the amount of psychological activity to be found in each country.

EGYPT¹

There are approximately 40 persons in Egypt who are engaged full time in psychological work. More than 30 of these are engaged in academic work, and most of the remainder are employed by the government. Two psychologists were cabinet ministers at the time this survey was made (2).

Two generalizations can be made about psychology in Egypt. First, the influence of British universities is very strong. Second, there is a close alliance between psychology and education. The second fact is probably related to the first. There are many individuals, of course, to whom these statements would not apply, and there is some evidence that British influence is less strong today than five years ago. Nevertheless, more than half of the psychol-

¹ We wish to express our appreciation to Dr. Hanna Rizk of the American University at Cairo for invaluable assistance in collecting data on Egyptian psychologists.

ogists in Egypt today received their advanced degrees from British institutions, and British journals rank next to Egyptian journals as vehicles for publication of research data. Only two persons reported having published in American journals, and two reported having published in France.

Nearly all of Egypt's psychologists are to be found in Cairo and Alexandria. A majority of the psychologists engaged in academic work are in one or another of the Institutes of Education. The Higher Institute of Education for Men, now a part of Cairo's Ibrahim University, has 11 psychologists on its staff. Four psychologists are with the Higher Institute of Education for Women in Cairo, two are with the Higher Institute of Education in Alexandria, and at least six more are on the staffs of the various normal schools. As might be expected from their institutional affiliations, more psychologists publish in the *Egyptian Journal of Education* than in the *Egyptian Journal of Psychology*. Indeed, this latter publication has recently changed from a quarterly to an annual.

Dean of the Institute of Higher Education for Men, Cairo, is Dr. Abdel Aziz El Koussy, whose graduate training was taken at Birmingham under Valentine and at London under Spearman and Burt. Specialists in clinical and developmental psychology, statistics, tests, and learning are on the Institute's faculty. These specialists hold Ph.D. degrees (or the equivalent) from Cambridge, London, Reading, and Leeds. All undergraduate students at the Institute take at least one course in psychology, and students in the newly organized graduate school take approximately half of their course work in psychology. In 1952, 81 persons

were taking graduate work, and 24 had passed the master's examination. Of those who had passed, about one-third were actively working on a master's thesis. The Institute is authorized to award both the M.A. and Ph.D. in educational psychology. At this writing no Ph.D. had been awarded.

The laboratory at the Institute is similar to that in many American universities. Several rooms are available; each is furnished with worktables and cabinets filled with apparatus. The apparatus include chronoscopes, mazes, mirror-drawing devices, galvanometers, kymographs, manual dexterity boards, instruments designed to measure mechanical ability, intelligence tests, and a large variety of paper-and-pencil tests. Undergraduate students who take experimental psychology are expected to perform four experiments from a list of experiments provided by the instructor. Candidates for the master's degree perform 20 experiments. English and American books serve as references. Foreign journals are not readily available to the students.

The Institute offers some work in clinical psychology and there is a clinic connected with it. The demands for the services of the clinic far exceed its capacity. Patients are limited to children and students. Three professors devote full time to the clinic, and a neurologist, a psychiatrist, and a social worker are available on a part-time basis. Dr. Ishak Ramzy is in charge. He holds a Ph.D. from London and was trained in psychoanalysis at the London Institute of Psychoanalysis, where he studied with Anna Freud.

There is an active testing movement in Egypt, and the Institute is one of the foci of this activity. Adaptations of the Binet and of

several of the Spearman tests have been in use for years. Pioneers and leaders in this area were Dr. Koussy and Dr. Ismael Kabbani, formerly with the Institute and now Minister of Education. At the present time the Army is using intelligence tests and other aptitude tests.

The Faculty of Arts of Ibrahim University, Cairo, offers a *license* in psychology which is roughly equivalent to the American master of arts degree. Chairman of the department is Dr. Mustafa Ziwer who holds both the M.D. and *Dr. es Let.* from the University of Paris. The latter degree was earned under Piéron. Although the department is small, an imposing array of courses is offered by drawing on the staff of the Institute of Higher Education and on the staff of the Faculty of Arts, University of Cairo. The program requires four years beyond the *baccalaureate deuxième partie*, and the first class to have completed this entire new program was graduated in 1954. It is expected that the graduates will be employed by the Ministries of Education and Social Welfare. Work toward a Ph.D. degree is to be offered in the future.

At the University of Cairo, the psychology department is a part of the philosophy section of the Faculty of Arts. Students working toward a *license* in philosophy take one psychology course in each of the last three years. Chairman of the department is Dr. Yusuf Mourad, a graduate of the University of Paris. He is co-editor, with Dr. Ziwer, of the *Egyptian Journal of Psychology*, and has contributed to it extensively. For his advanced courses he assigns works in French by Guillaume, Piéron, Wallon, and others, and French translations of such American authors as Woodworth, Carmichael, and Krech and Crutchfield. Courses

in experimental psychology are offered at Fouad by Dr. Muhammad Nagati, who did his graduate work at Yale University.

The Higher Institute of Education for Women, Cairo, and the Alexandria Higher Institute of Education for Men both offer work in educational psychology. Each of these institutes has at least three full-time professors in this field. The American University at Cairo offers courses in psychology as part of its education and social studies curricula. Psychological training is also a part of the program given by the Cairo Institute of Social Work.

Al Azhar, Cairo, one of the world's oldest universities, is traditionally devoted primarily to theological training. There are today three sections at the higher level: theology, Islamic law, and Arabic language. The section devoted to the teaching of the Arabic language now offers courses in general and educational psychology, and the theology section offers one course in psychology.

LEBANON²

Since the time of the Crusades, and especially for the past hundred years, Lebanon has had close cultural ties with the West. At present one finds in the book stores of its cities large numbers of Western books, including scientific and pseudoscientific works on psychology. Works on psychoanalysis are so popular that a leading Beirut newspaper ran in 1953 a series of editorials denouncing the "efforts of Western psychologists" to turn attention from social ills to problems of infantile frustration. There is even one diabetics practitioner in Beirut who re-

² We wish to express our appreciation to Miss Suad Khalluf for assistance in obtaining information about psychology in Lebanon.

ceived training from Hubbard himself.

There are three universities in Lebanon, and these universities are the foci of psychological activity in the country. *The Centre d'Etudes Supérieures Français de Beyrouth* is a branch of the Arts and Sciences Faculty of the University of Lyon. It offers a *license* degree which requires that the student complete five certificates. The five can be obtained in two or three years. Of the many certificates which may be chosen, three are in the field of psychology: general, social, and educational. Each certificate requires from six to eight hours per week for an academic year. The professors of the psychology courses hold chairs of philosophy. They place great emphasis on preparation for the annual examinations, which are issued from Lyon. They report, in familiar terms, that their teaching keeps them too busy for research.

The first psychological work published in the Arabic language in modern times seems to have been *'Al durūs al 'awaliyah fil falsafah al 'qliyah* (First Lessons in Mental Philosophy). It was written in 1874 by the American missionary Daniel Bliss, who was founder and first president of the Syrian Protestant College (now the American University of Beirut). Today the psychology department of the American University resembles that of the liberal arts college in the United States. Courses are taught in English, and American textbooks are used. An eight-hour course in experimental and a three-hour course in statistics are required of all majors. The university emphasizes research, and for more than a decade the department has concentrated on research in social psychology. This tradition was begun by S. C. Dodd, now at the Univer-

sity of Washington, who was for many years chairman of the sociology and psychology departments. Closely allied with the American University is the Beirut College for Women. The College stresses work in developmental, guidance, and educational psychology.

The Lebanese National University held its first classes in 1950, and the psychology program there is in an embryonic stage. Some courses are offered now and it is planned to offer a certificate in psychology in 1956. The organization and content of the courses reflect the influence of the French university and, to a lesser extent, of the American University.

SYRIA

Syrian psychology centers in the Syrian State University located in Damascus. The university is quite new. Although the Schools of Medicine and Law date back to 1923, it was only in 1946 that the Schools of Arts, Science, and Engineering, and the Higher Normal School, were joined with the older faculties (1). Some of the buildings occupied by the university are former Ottoman and French army barracks. Psychology courses are offered in both the philosophy faculty of the College of Arts and Sciences and in the education department of the Higher Normal School. In order to obtain a *license* in the philosophy faculty, a student must study philosophy, psychology, and sociology. If he receives the degree, he is entitled to teach these subjects in secondary schools. The secondary schools, like those of France, carry the students to approximately the American sophomore level. Of the four professors on the philosophy faculty, two are philosophers and the other two are psychologists on loan from the Higher Normal School.

The education department is quite active in psychology, and the members of its faculty offer a wide variety of courses. Chairman of the department is Dr. Fakhir Akil, who received B.A. and M.A. degrees from the American University of Beirut, and a Ph.D. from London. His doctoral work was done under Drs. Philpott and Burt. Each of the other full-time professors holds a *license* from Egypt and is completing a doctorate in France. Consequently, Syrian psychology is a mixture of American, English, and French approaches. The professors are writing textbooks in Arabic, and they assign collateral reading in both French and American works. Woodworth's *Experimental Psychology* and Morgan's *Physiological Psychology* are used in their French editions. The laboratory is well equipped with pursuit meters, mirror-drawing apparatus, form boards, intelligence test materials, personality test materials (including Rorschach, TAT, Spearman Abilities, and Szondi), and many other devices. We were impressed to learn at the time of our visit to the laboratory that the annual budget for new laboratory equipment was approximately \$4,500.

Students achieving satisfactory scores on entrance examinations may study at the Higher Normal School at government expense. If they wish to accept the subsidy, they sign a contract with the Ministry of Education whereby they agree to work for a specific number of years for that ministry upon completion of their work. Moreover, the top student each year is sent abroad to work for a doctor's degree. Partly as a result of this arrangement, there is strong student interest in educational psychology. The trend seems to be away from philosophical psychology, in the French tradition, toward

educational psychology of the British and American variety. This trend may be a function of Dr. Akil's leadership, and it is too early to say whether it will continue when students now studying abroad return to Syria. Because of currency problems, transportation costs, and other factors, more Syrians are studying in France than in Britain and America together (4). In Syria as in Egypt, there is more psychology work in the education department than in the philosophy faculty.

IRAQ

A visit to modern Baghdad gives one the impression that there is a lively curiosity about psychology in literate circles. Bookstores stock and presumably sell many English-language psychology textbooks, even though these textbooks are not assigned in any Iraqi university. Woodworth's *Experimental Psychology*, for example, is apparently read by some Iraqis out of sheer intellectual curiosity!

Although many of the professors are American educated, education in Iraq is dominated by persons trained in Britain. As in Egypt, educational psychology is taught in both urban and rural teacher training institutions. Teachers of these courses are persons who were trained primarily in the field of education.

There is a department of education and psychology in the Higher Teacher's Training College, located in Baghdad. Several courses are offered in psychology, and there are plans for expansion. The principal members of the staff are Dr. Ibrahim Muhyi, who completed his doctorate at Columbia University, and Dr. Abdul Aziz Al-Bassam, whose graduate study was at University College, London. Each of these professors also offers courses at Baghdad's Col-

lege of Arts and Sciences and at the Queen Alliya College for Women. This last-named school offers psychology in the department of social work. As far as we know, there is no psychological laboratory in Iraq, and little psychological research.

THE PERSIAN GULF AREA

There is hardly any psychological activity on the West (Arab) side of the Persian Gulf. In Jordan there are two teacher training institutions where Egyptian-trained educators offer elementary work in psychology. There are no trained psychologists in Bahrein or Kuwait, and none in the Arabian peninsula except the Americans who are employed by the Arabia-American Oil Company.

CONCLUSION

From this brief survey it can be seen that psychology in the Arab Near East is essentially academic. Where the French academic tradition prevails, psychology is closely allied with philosophy. Where British academic tradition prevails, psychology is allied with education, and there is a strong emphasis on testing and measurement. There is some concern with research in Egypt and Lebanon, but this activity is subordi-

nate throughout the area to translation of important Western books and adaptation to local use of Western tests.

In his discussion of "the sociology of psychology" Sanford (3, p. 85) has suggested that dictatorial regimes are antipsychological. Yet the military regimes in both Egypt and Syria seem eager to utilize whatever contributions psychology can make to the perceived social problems. Interest in psychology on the part of these governments exceeds that of the parliamentary governments which preceded them. This does not demonstrate that *antidemocratic* trends favor psychology, for the political form is not always the best index of the essential nature of a society.

The prognosis is favorable for the growth of psychology in this part of Western Asia. Changes in long-established social institutions, increased urbanization, increased literacy, and rising per capita income all point to increased utilization of psychological understanding. There is some suspicion that psychology is an instrument of Western imperialism, but there is reason to hope that this suspicion will be overcome as increasing numbers of Arabs become psychologists.

APPENDIX

BIBLIOGRAPHY OF SELECTED TITLES

There is no published bibliography of recent books in the Arabic language which deals with psychological topics. The best source of information about such books is probably *'Al Sijil al Thakafi* (Cultural Register), published annually by the Egyptian Ministry of Education in Cairo. The register contains information about publications in many fields, including psychology and education.

We have listed below some books which we consider to be representative of the best works on psychology published in Arabic from 1945 to 1953. The list is representative rather than inclusive.

1. 'Al 'Ahwani, 'Ahmad Fu'ad. [Outline of psychology.] Cairo: Amiriya Press, 1948.
2. Al Athama, Wafic. [Modern psychology.] Damascus: Hashimiyah Press, 1950.
3. Al Draubi, Sami, & 'Al Jamali, Hafiz.

[Psychology.] Damascus: Dar-'al-Yaqtha Press, 1949.

4. Al Khawli, William. [The integrated personality.] Cairo: Mukhaymar Press, 1948.
5. Al Qusi, Abdul Aziz. [Foundations of

- mental health.] (2nd Ed.) Cairo: Al Nahda Press, 1952.
6. Al Qusi, Abdul Aziz. [Foundations of Psychology.] Cairo: Al Nahda Library, 1950.
 7. Al Qusi, Abdul Aziz, *et al.* [Language and thought.] Cairo: Amiriyah Press, 1948.
 8. Al Qusi, Abdul Aziz. [Statistics in education and psychology.] Cairo: Al Nahda Library, 1949.
 9. Al Mulayhi, Abdul Mun'im. [Psychological development.] Cairo: Misr Library, 1950.
 10. Al Tawil, Tawfic. [Dreams.] Cairo: Al Adab Press, 1955.
 11. 'Akil, Fakhir. [Psychology and its application to education.] Damascus: Maktabat Al'uloom Wal'Adab, 1945.
 12. Barakat, Mohamad Khalifah. [Analysis of personality.] Cairo: Misr Library, 1951.
 13. Barakat, Mohamad Khalifah. [Psychological clinics.] Cairo: Misr Library, 1952.
 14. Jirjis, Sabri. [The problem of psychopathic behavior.] Cairo: Dar Al Ma'rif, 1952.
 15. Khalaf Allah Muhammed. [A psychological approach to the study of literature and literary criticism.] Cairo: Lajnat Al ta'leaf Wal tarjamat Wal Nashr, 1947.
 16. Mitri, 'Amin. [The feeble-minded.] Alexandria: Dar Nashr 'Al Thaqafah, 1948.
 17. Murad, Yusuf. [Elements of general psychology.] Cairo: Dar Al Ma'rif Press, 1948.
 18. Najati, Mohamad 'Uthman. [Military psychology.] Cairo: Dar Al Ma'rif, 1953.
 19. Ramzi, Ishaq. [Individual psychology.] Cairo: Dar Al Ma'rif Press, 1946.
 20. Saliba, Jamil. [Psychology.] Damascus: 'Ibn Zaydoun Press, 1948.
 21. Salih, 'Ahmad. [Educational psychology.] Cairo: Al Nahda Library, 1951.
 22. Subay'i, 'Adnan. [A survey of child psychology.] Damascus: 'Arafah Library, 1952.

REFERENCES

1. MATTHEWS, R., & AKRAWI, M. *Education in Arab countries of the Near East.* Washington: American Council on Education, 1949.
2. PROTHRO, E. T. Psychologists in high posts in government. *Amer. Psychologist*, 1953, 10, 597.
3. SANFORD, F. H. Toward a sociology of psychology. *Amer. Psychologist*, 1952, 7, 83-85.
4. WILLIAMS, H. H. *Syrians studying abroad.* New York: Institute of International Education, 1952.

Received May 15, 1954.

THE ROLE OF MOTIVATION IN VERBAL LEARNING AND PERFORMANCE¹

I. E. FARBER

State University of Iowa

The factors affecting performance in learning situations are customarily divided into two classes—the associative and the nonassociative. The former have been the subject of most research in the area of verbal behavior, and the information concerning them constitutes by far the better part, in quality as well as quantity, of what is known about this kind of behavior. In recent years, however, there has been an increasing interest in those factors that are commonly supposed to have nonassociative properties, particularly those subsumed under the heading of "motivation."

Although the term "motivational" is often used in reference to non-associative factors, conventional psychological usage imputes to motives an associative aspect also. Melton, for instance, considers motives as having three functions: the energizing, the selective, and the directing. According to his view, the energizing function "... may be most simply conceptualized as a result of the lowering of the *general* (italics added) reaction threshold by the motive" (30, p. 672); the selec-

tive function refers to the reinforcing effects of the achievement of goals relevant to the motive; and the directing function is associative or cognitive, considering motives, as part of the stimulus complex, as having a given habit strength in respect to various specific responses. According to Shaffer (41), a motive involves both "drive" and "mechanism," the former term denoting a variable that arouses mass activity, and the latter, an acquired response tendency. Hull (23), in his discussion of primary motivation, supposed that while "... all drives alike are able to sensitize all habits" (p. 247), "associated with every drive (*D*) is a characteristic drive stimulus (*S_D*)..." (p. 253), which gives specificity and direction to behavior. And Dollard and Miller (13), in discussing the functional properties of stimuli, refer to their drive aspect—that which impels action generally, and their cue aspect—that which leads to differential response.

The supposition that motivational variables have steering (associative) properties as well as energizing and reinforcing (drive) properties does not, of course, imply that the two kinds of properties are identical. Indeed, the virtue of the foregoing analyses of the concept of motivation lies in their separation of these functions. The associative function of a motive or any other variable is identified in terms of its tendency to elicit or facilitate a limited class of responses only. The drive function of a variable is demonstrated if: (a) its presence energizes or intensifies indiscriminately all reaction tend-

¹ The present paper is a revision of a paper entitled "Motivational factors in verbal learning," presented May 8, 1953, in the *Symposium on Psychology of Learning Basic to Military Training Problems*, sponsored by the Panel on Training and Training Devices, Research and Development Board. It was prepared as part of a project on the influence of motivation on performance in learning, under Contract N9 onr-93802, Project NR 154-107, between the State University of Iowa and the Office of Naval Research. The helpful criticisms of this manuscript by Drs. K. W. Spence and J. S. Brown are acknowledged with thanks.

encies existing in a given situation; and/or (b) its elimination or reduction in magnitude is reinforcing, i.e., leads to the increased probability of recurrence, in the same situation, of the responses preceding its modification. Motives, it is presumed, have both functions, and if a variable clearly does not have both, its status as a motive is questionable.

But, as Brown has recently pointed out, "an examination of contemporary discussions of motivation suggests that one of the major sources of misunderstanding is the failure of most writers to distinguish clearly between drives or motives, on the one hand, and habits or reaction tendencies on the other" (5, p. 2). According to his view, all instances of directed behavior are attributable, not to drives or motives as such, but rather to the capacity of stimulus cues, including those accompanying drives states, to elicit particular reactions.²

Brown's restriction of the use of the terms "motive" and "motivational" to refer to dynamogenic functions alone accords with Hull's

recent formulations, as discussed by Spence (47, 48), and is believed to be in the best interests of terminological clarity. However, these terms will be used in the present paper in their more traditional (and ambiguous) meaning, to refer to states or processes having both directing and dynamogenic components. The directive aspects of motives will be referred to variously as "associative," "cognitive," or "habitual," whereas the term "drive" will be used to refer to the nondirective, energizing and reinforcing aspects, in accordance with Hull's usage of the concept *D*.

Since certain variables not usually classified as motivational, e.g., ordinary environmental stimuli, may have strong associative connections even though they have little or no drive value, it is assumed that it is their drive function, basically, that distinguishes motivational variables from other kinds. Thus, the mere demonstration that a given variable has directive characteristics is not enough to qualify it as motivational. Furthermore, even if a variable qualifies as motivational because of its evident energizing or reinforcing effects under some circumstances, its effects upon behavior in other circumstances may depend primarily upon its associative properties, i.e., its tendency to elicit particular responses. On the other hand, the fact that a variable has associative properties does not preclude the possibility of its having drive properties as well.

The implications of this general view are here considered in respect to the interpretation of the role of certain variables, commonly regarded as motivational, in verbal learning and performance. Although the discussion is somewhat arbitrarily limited to only a few such vari-

² Postman has suggested that the notion of a nondirective drive may make the concept of drive expendable. Thus, he proposes that the explanation of variability in behavior, despite the constancy of stimulus situation and habit strength, might well be sought in the "changing patterns and intensities of drive stimuli" (37, p. 57). This sort of formulation, in point of fact, bears a close resemblance to that of Dollard and Miller (13), who define a drive as a strong stimulus. It does not, however, as Postman further suggests, endanger the Hullian distinction between sH_R and sE_R , nor for that matter, the somewhat similar Tolmanian distinction between cognition and performance. Hull's *D* (as opposed to S_D) is nondirective; a concept like drive would presumably still be needed to explain at least some kinds of reinforcement phenomena; and intensity of stimulation, in Hull's last formulations, is itself considered a motivational variable multiplying habit strength (cf. 47, 48).

ables in relation to human verbal behavior, it is believed that these variables are not unrepresentative of motivational variables in general, and also, that the considerations raised concerning the analysis of their effects apply as well to nonverbal behavior, in both animals and men (cf. 15).

Experimentally Manipulated (Environmentally Defined) Motives

Electric shock. In a large number of investigations of the effects of motivation, the state of *S* is more or less under the control of the investigator. These "experimentally manipulated" states have, for the most part, been acquired. However, one primary motivational state whose effect upon verbal learning and performance has received considerable attention is that induced by electric shock.

To say that the effects of electric shock are not very well understood is an understatement. Nevertheless, an examination of studies involving its use is instructive, since it indicates many of the considerations that must attend analyses of so-called motivational variables in general. First of all, it seems clear that electric shocks qualify as motivational by virtue of their drive-producing effects. The role of electric shock as a reinforcer is well known, and its general energizing effect has been demonstrated in a number of studies (e.g., 1, 35, 42, 51). But these are by no means its only functions. For instance, electric shocks may be informative. Thus, under certain conditions, particularly when appropriate verbal instructions are given, electric shock may lead to improved performance, whether administered for wrong responses or for right ones (28). It is probably gratuitous, and even misleading, in these instances,

to attribute the effects of the shock solely to its drive characteristics. Since shock has not only drive value, but also cue value, its associative or cognitive aspects must be taken into account.

Even in those instances in which shock is noninformative, i.e., not related in any systematic way to correct or incorrect responses, it is necessary to consider its response-producing properties. The fact that shock tends to elicit withdrawal, or vocalization, or muscular contraction in rats or human beings, may be a consequence, not of its drive characteristics, but rather, of its association with either learned or innate response tendencies. Since shock is an intense stimulus, capable of eliciting a variety of responses, it is not surprising that in certain situations it may elicit responses that are incompatible with those elicited by other stimulus components of the situation. Certain kinds of verbal performance are perhaps particularly susceptible to disruption (interference) by shock, though there seem to have been no investigations specifically designed to discover whether verbal behavior is more susceptible in this respect than nonverbal behavior of various kinds, particularly when other factors, such as level of practice, are held constant.

Although the disrupting effect of shock, when it occurs, is often attributed to the disorganizing influence of emotion, it seems unnecessary, at least in some situations, to attribute the interference to anything more than the introduction of an intense, extraneous stimulus which produces strong competing responses. This interpretation may also apply to other kinds of prepotent stimuli that have a distracting or disrupting effect, e.g., loud noises, loss of support, or simply talking to *S*.

It must be noted that some kinds of verbal behavior, e.g., swearing, are likely to be elicited by shock or other noxious stimuli as a result of learning. If such responses are considered "correct" in verbal learning situations, shock might be expected to facilitate performance. Depending, then, upon the particular responses that have been acquired and the particular requirements of the test situation, shock might be either beneficial or detrimental. Unfortunately, little is known concerning the relative strength of various kinds of verbal response to shock. It does not seem improbable, however, that at least some of the effects of shock are attributable to learned associations rather than to its drive value as such.

Hunger. A somewhat more confident assessment of the role of stimulus cues may be made in relation to certain studies of the effects of the primary drive of hunger. Thus, there is evidence, not without its exceptions, to be sure, that increasing hunger increases the probability of occurrence of verbal responses referring to food. When this does occur, as both Brown (5) and Postman (36) have pointed out, it is most easily explained on the supposition that responses relating to food are more strongly associated with hunger stimuli than are most other responses.

This is not to say there is no evidence that hunger, defined in terms of food deprivation, functions as a drive. Studies showing that the receipt of food following deprivation is reinforcing, or that the presence of hunger has a generally enhancing effect upon response tendencies, attest to its drive properties. But when the responses elicited or intensified in the presence of a drive are "need-related," i.e., when they are responses that have been conditioned to the

cues characteristic of a given drive, their explanation obviously requires no appeal to drive *qua* drive. Studies showing that the frequency of verbal responses relating to food or eating in the presence of impoverished or ambiguous verbal or visual stimuli increases as a function of hunger are interesting and important. But they appear to deal, for the most part, with the associative rather than the drive properties of hunger.

Fear (anxiety) resulting from noxious stimulation. Many verbal learning studies involving electric shock are designed to investigate its effects upon subsequent performance under nonshock conditions. It is not improbable that behavior under these conditions is in part a function of purely perseverative effects of pain, or what a number of investigators (e.g., 1, 42, 43) have termed "emotionality." However, the motivational state usually considered in this connection is that presumably aroused by environmental cues associated with the shock. The state thus aroused is an acquired one, namely, fear or anxiety. The demonstration that fear, so defined, has nonassociative properties rests primarily, thus far, upon animal experiments indicating that the reduction of fear is reinforcing (31), or that its occurrence intensifies ongoing responses that are not themselves produced by the fear (7).

As in the case of hunger, many of the experimental studies of the effects of fear on human behavior involve the learning or retention of motive-related materials, i.e., particular words or other items that had previously appeared in close temporal contiguity with a noxious stimulus. As a result of conditioning, such items might be expected to elicit not only fear, but also some of the other responses previously elicited

by the shock. But in addition, as Dollard and Miller (13) point out, fear itself brings with it a number of response tendencies, either learned or unlearned. (This holds true, of course, even though the fear is produced by stimuli other than those comprising the material to be learned or recalled.)

The empirical evidence concerning the effects of fear on verbal performance has not been altogether consistent. Under certain conditions the production of fear seems to facilitate the recall or recognition of verbal materials (12, 34). More often, however, it seems to elicit responses that successfully compete with verbal behavior. Considered in this manner, fear-produced responses might well be classed among those more prosaic variables usually considered as determinants of associative interference or facilitation. We have, as yet, very little information concerning the kinds of response tendencies ordinarily produced by fear, or the way in which they interact with other response tendencies in specific situations. There is reason to believe that fear may have particularly intense stimulus properties, and, other things being equal, may therefore elicit unusually strong responses, of a sort not ordinarily observed in laboratory studies of forgetting (cf. 40). But this does not in itself necessitate an explanation of its effects in other than associative terms. Again, it must be emphasized that this is not to deny that fear acts as a drive. It does deny that disruption (or for that matter, facilitation) of behavior by fear is in itself an adequate basis for the supposition that it has drive properties (15).

Anxiety aroused by socially tabooed materials. Somewhat related to investigations using items associated with noxious stimulation are the

studies involving socially tabooed verbal materials. In these studies it is supposed that certain words or phrases arouse an acquired anxiety drive, not because of their association with punishment in the experimental situation, but because of previously experienced social punishments, administered either for the overt usage of the verbal expression itself, for behavior labeled by the verbal expression, or for behavior associated with stimuli thus labeled. Anxiety aroused in this manner is sometimes regarded as more similar to the sort of anxiety considered by clinicians as a determinant of repression than that produced by electric shock (40).

In many of the studies in which it has been found that the learning, retention, or recognition of socially tabooed or negatively valued items is impaired, the interpretation is often in terms of unconsciously motivated defense mechanisms. This sort of interpretation has been seriously challenged on two grounds: first, that the mechanism involved is typically far from unconscious; and secondly, that it primarily involves habitual rather than purely motivational mechanisms.

The first objection stresses the possibility that the overt expression of socially unacceptable responses is at least sometimes deliberately and consciously inhibited (e.g., 21, 36, 57). The second objection rests on the finding by Howes and Solomon (22) that the visual recognition thresholds for words vary with their relative frequency of usage, as determined by the Thorndike-Lorge word counts. By taking account of the large difference in the mean frequencies of usage of the taboo and neutral words used in an investigation by McGinnies (29), Howes and Solomon (21) were easily able

to account for McGinnies' finding that the taboo words had the higher thresholds. Postman also has adopted the view that high thresholds for negatively valued words are explicable in terms of their relative habit strengths. Pointing out that the report of such words may be inhibited by the tendency to give more frequently used neutral words that have elements in common with the taboo words, he has concluded that

... it is not necessary to ascribe selective functions to the immediate motivational conditions of the perceiver. Defensive and/or vigilant thresholds as well as systematic pre-recognition hypotheses may be consequences of the relative strengths of response dispositions aroused by specific stimuli (36, p. 80).

This interpretation, it should be noted, is not inconsistent with the view that socially disapproved words may arouse anxiety, and that the anxiety acts as a drive. It is not improbable that the low frequency of usage of these words is itself due to their evocation of anxiety and to the reinforcement of their suppression resulting from anxiety reduction.³

Failure defined by verbal instructions. One of the most widely used techniques for inducing so-called motivational changes in verbal learning situations is the use of instructions by *E* indicating that *S* has failed. As Lazarus, Deese, and Osler

(26) have observed, a distinction must be made between the purely motivational (drive ?) effects of failure instructions and their informative or distracting effects. Thus, the problems of distinguishing between the drive and the associative components of motivational variables must be considered in relation to failure experiences also.

Experimental studies have usually tended to show that when failure instructions are not applied to specific responses, but are used, instead, to characterize general performance, subsequent performance in verbal learning situations is impaired. As a result of such observations, it is generally believed that statements expressing approval of a learner's performance are more beneficial to his subsequent performance than statements indicating disapproval. At the same time, however, it is also rather generally believed that excessive praise may decrease motivation and thus impair performance, whereas some kinds or degrees of censure may increase motivation and thus improve performance. This indecision with respect to the effects of failure characterizes treatments of the effects of frustrating events in general. For instance, Child and Waterhouse write:

Indeed, the conflict appears strikingly in some general textbooks in psychology. In a chapter on thinking and reasoning frustration is viewed as the condition for more organized behavior, and in a chapter on emotion it is viewed as the condition for less organized behavior. The failure to use a common term such as frustration in the two chapters apparently permits the contradiction to go unnoticed (10, p. 127).

In their subsequent discussion, these writers have presented a notably useful analysis of the manner in which the responses produced either directly or indirectly by the frustrating event might either im-

³ This account is clearly neutral in respect to the question of the usefulness of the concept of "perceptual defense." It affirms the possibility that anxiety cues may produce responses that successfully compete with correct recognition, and is altogether in accord with the view that there may be marked individual differences in the amount of anxiety aroused by various cues and/or the nature of the responses elicited by this anxiety. Furthermore, following Dollard and Miller (13), it freely admits the possibility that the individuals concerned may sometimes be unconscious of the cues eliciting the anxiety.

prove or impair performance. They suppose that prevention of the dominant response in the hierarchy may lead to its extinction, with the result that initially weaker responses are now evoked; or the frustrating event may modify the situation in such a way that some new response is elicited, either by some changed element of the situation or by the frustrating event itself. These possible effects of frustrating events, it should be noted, relate exclusively to the modification of cues and/or to the relative strengths of habits associated with these cues.

This emphasis by Child and Waterhouse upon the associative aspects of frustration extends also to their treatment of the possible drive-producing consequences of frustrating events. Thus, they suppose that "... frustration may operate to increase the motivation supporting the goal-oriented activity and thereby to improve the quality of performance" (10, p. 136), or conversely, to produce a drive state of intense general emotion which impairs performance either "... because the emotional responses themselves are to some extent incompatible with the ongoing instrumental activity" (10, p. 137), or because the emotion produces other incompatible responses. It would appear, then, that the only properties explicitly attributed by these writers to the motivational states that might be produced by frustrating events are directive.

They suggest, correctly, that the foregoing treatment has something in common with the "emotional" interpretation of frustration behavior by Brown and Farber (6). But it seems, in general, to fall more clearly within the class of "nonemotional" interpretations, as discussed by the latter writers, since it deals with the possible behavioral consequences of

frustration without appealing to the concept of a frustration-produced drive having nonassociative properties. As a matter of fact, until recently there has not been much evidence of a positive nature to support the widespread notion that frustration produces (or acts as) a drive. Consequently, it has been doubtful whether, from the present point of view, frustrating events qualify as motivational variables (cf. 6). However, recent studies by Amsel and Roussel (2) and by Lambert and Solomon (25) indicate that at least some kinds of frustrating events affect animal behavior in the manner characteristic of a drive.

Whether failure, defined in terms of verbal instructions, has a comparable effect upon verbal behavior still remains to be demonstrated. Unfortunately, under most conditions of verbal learning, there is no easy way to separate associative from possible drive effects of supposedly motivational variables. Contrary to general belief, the impairment of performance by failure does not necessarily demonstrate its associative effects, nor does the improvement of performance by failure demonstrate its drive effects. Just as the associative mechanisms attending failure may either benefit or harm performance, so might an increase in drive, in and of itself, affect behavior either favorably or adversely, depending on the specific nature of the task and the experimental conditions involved (16, 48). Since few, if any, studies of the effects of failure have thus far been designed in a manner that permits an evaluation of its drive properties, the motivational status of this particular variable still seems somewhat doubtful.⁴

⁴ Evidence concerning the reinforcing effects of escape from or avoidance of failure must not be neglected, of course, in considering this

Differentially Selected (Response-Defined) Motives

Logic of response-defined hypothetical constructs. The variables discussed so far have consisted of directly manipulable environmental conditions. Very often, however, motivational states are defined in terms of performance measures of one sort or another. In order to observe the effects of variations in motivation, thus defined, one must differentiate or select individuals who vary in terms of the particular response index chosen. If the response measure does reflect degree of drive, and if one has a theory concerning the way in which such variations influence behavior, information concerning this response index for individual Ss should enable one, in specific situations, to predict their behavior more accurately than one could do otherwise.

As Spence has pointed out (45), the usefulness of response-inferred theoretical constructs is in some ways limited, and, in principle at least, it should be possible, given sufficient information concerning antecedent conditions, both historical and contemporaneous, to explain or predict the very behavior used as the index of the hypothetical state under consideration. However, when information concerning these antecedent conditions is meager or unavailable, the use of responses as indices of intervening variables is clearly advantageous. The importance of response-inferred constructs has been particularly stressed by those psychologists who insist that behavior cannot be predicted at all without ascertaining

problem. It seems probable that failure or anticipation of failure can be profitably conceptualized in terms of a state resembling anxiety, which, when defined in certain ways, seems clearly to qualify as a motive (15).

the "meaning," or S's "perception" of a situation. In practice, they use certain responses, usually verbal, as indices of a "psychological environment," or "phenomenological field," or some other hypothetical structure, to which they assign a primary explanatory role. From an S-R point of view there can be no quarrel with this procedure, provided that the responses from which the hypothetical states are inferred are explicitly and unambiguously identified. However, as Bergmann observes, it is "... the business of science to ascertain which objective factors in the past and present states of the organism and its environment account for the difference in response ..." (4, p. 133), including, presumably, the response(s) from which one infers the "meaning" for S of any given situation.

Of course, statements of relations between one kind of behavior and another, which Spence (45, 46) calls "R-R laws," need not involve any theory about motivation, or, for that matter, any other hypothetical state. But if a given response measure is labeled an index of motivation, and if this labeling is to serve any useful purpose, one's theory concerning the relation of motivation to behavior ought to be made explicit. Otherwise, there is no justification for considering the response in question to be a measure of motivation rather than anything else.

It is instructive, from this point of view, to examine the many response measures that have been proposed as indices of motivational states. When this is done, one finds a very meager basis indeed for supposing that the majority reflect anything but associative tendencies. Although such findings do not in any way minimize the importance of these instruments or the empirical relations established in connection with them, they empha-

size the necessity for formulating adequate criteria of motivation.

Allport-Vernon Study of Values. The problem of the motivational status of response-defined variables has received a good deal of attention recently in the specific case of performance on the Allport-Vernon Study of Values. As a result of the able analyses by Solomon and Howes (44) and Postman (36), it seems evident that variations in scores on this particular instrument are much more reasonably interpreted in terms of variations in the habit strength of certain verbal responses than in terms of variations in drive factors per se. Studies by these investigators have demonstrated empirically that the relative frequency of usage of words in the various value areas comprising the Allport-Vernon scale is the most important determinant of recognition thresholds for words associated with these value areas.⁵ But on theoretical grounds alone it could be concluded that a response index that does nothing but predict the relative strength of other specific responses, and particularly those highly related to the items comprising the index, does not qualify as a measure of drive. There is no question that the Allport-Vernon Test measures something; but, from the present point of view, there is no reason to suppose that the "something" is a motive.

Achievement imagery. Recently, McClelland, Atkinson, Clark, and Lowell (27) have published an important analysis of motivation, centered about their experimental investigations and those of their colleagues and students, of achievement

imagery, as measured by the contents of stories given in connection with TAT-like pictures. The results of these investigations have indicated that Ss who vary in their achievement imagery scores tend to differ, under certain conditions, in various other kinds of performance, including verbal. The predictive usefulness of this behavior index is clearly attested to by such findings, but it is not entirely certain that they demonstrate that the test reflects motivational characteristics.

Thus, the obtained relations between achievement imagery scores on the one hand, and visual recognition thresholds for achievement-related words or the frequency of usage in a written essay of certain terms (see 27, p. 252) on the other, seem to be precisely those that would be expected on the assumption that variations in achievement scores reflect variations in the strengths of certain verbal *habits*. In other words, the interpretation by Solomon and Howes and by Postman of the behavioral correlates of scores on the Allport-Vernon Scale may apply with equal cogency to these results also.

The finding that increment of output with practice in a Scrambled Words Task was positively related to achievement scores is given particular weight by McClelland and his colleagues in arguing that scores on their test are measures of motivation. They point out that, for some theorists, "... the decisive criterion for determining whether a motive is involved in performance is whether or not it can produce learning. Since there is clear evidence of learning in the high π Achievement group, it can be argued that the π Achievement score is a measure of motivation" (27, p. 234). From the present point of view this reasoning is not very convincing. Although motivation is held

⁵ This does not mean that variations in the habit strength of the verbal responses might not themselves be attributable to variations in drive. This possibility, however, would require independent verification.

to be essential to learning, it does not follow from this assumption alone that, with other factors presumably held constant, better performance in learning situations is necessarily due to a higher drive level. Indeed, Hull's theory leads to the prediction that, under certain learning conditions, Ss with higher drive will perform more poorly than those with lower drive (16, 33, 38, 48, 54).

The demonstration that the amount of food-related content in Ss' stories tends to be positively related to number of hours of food deprivation has also been used by the Wesleyan group as evidence of the influence of drive factors upon the kind of imaginative productions from which they have derived their *n* Achievement scores. Since hunger, defined in terms of number of hours of food deprivation, qualifies as a drive by virtue of the demonstration, in other studies, of its enhancing and reinforcing properties, it is possible that the amount of food-related content in TAT stories reflects the drive aspect of hunger directly. But, as the discussion in an earlier section of this paper has indicated, an alternative interpretation in terms of the verbal habits associated with hunger is also clearly possible.

In general, the authors of *The Achievement Motive* assume that responses in fantasy situations are relatively unaffected by variations in habitual or cognitive factors:

That is, for most subjects putting thoughts into words or verbalizing is a highly overlearned response [and is therefore a "normal" habit, whose strength is substantially similar in most individuals]. Furthermore, in the fantasy situation no particular set of responses is supposed to be perceived as especially appropriate. Fantasy is a "free" response situation, provided the picture is not too structured. It might not be for a certain class of persons, for professional writers, for example, because they may have learned a particular set of responses to use in such a situation. . . .

But except for professional authors, individuals should have no particular set of verbal response tendencies which seem appropriate because of past experience with such situations (27, pp. 40-41).

Although it is somewhat difficult to evaluate the merits of this argument, since the meaning of the concept, "normal habit," is not altogether clear, there seems to be plenty of evidence that stimulus factors and specific past experiences play at least some role in determining TAT performance (e.g., 11, 14, 39). The factors determining the kind and amount of influence of associative variables upon performance in fantasy situations like that presented by the TAT are certainly not so well known as those determining the kinds of verbal performance traditionally studied in the laboratory. At present, therefore, the supposition that variations in performance on so-called "projective" tests do not reflect variations in associative factors, like the assumption that they do, appears to owe a greater debt to faith than to empirical fact.

Iowa Picture Interpretation Test (IPIT) achievement scores. Despite the foregoing demurrals, the extensiveness of the range of responses apparently affected by the achievement variable, as defined by the Wesleyan group, lends some credibility to the supposition that achievement imagery scores may indeed reflect strength of drive. Among the studies of verbal behavior, the most important, in this connection, are those that demonstrate a relation between achievement scores and other kinds of verbal behavior whose content has no obvious relation to that from which the achievement scores are themselves derived.

If achievement imagery scores are positively related to strength of drive, it could be predicted, on the

basis of Hull's (23) assumptions concerning drive combination, that the introduction of an experimental variable calculated to increase drive would have less effect on the behavior of Ss with high achievement scores than it would have, in general, on the behavior of those with low achievement scores (cf. 6 for a similar deduction in regard to frustration). Stated somewhat differently, a correlation might be expected under ordinary conditions between achievement imagery scores and performance. The addition of another drive variable, however, ought to reduce this correlation.

To test this hypothesis, Hurley (24) recently administered the IPIT to a large number of Ss. In this test, Ss are given a choice of four interpretations, or imaginative descriptions, of each of ten TAT pictures. One of the four interpretations of each picture involves achievement orientation, as judged by clinical psychologists on the basis of a definition taken from *The Achievement Motive*.⁶ The other three interpretations for each picture were designed, respectively, to indicate hostility, insecurity, and blandness.

When Ss were given instructions intended to be poorly motivating, a significant positive relation was found between achievement scores and the number of correct anticipations in a serial rote learning task. Under high motivational instructions, however, the correlation was insignificant.⁷ There was also some tendency for achievement scores to be positively

correlated with number of erroneous anticipations under the low motivational instructions. This result, though not statistically significant, was similar to a finding reported by Lowell (see 27).

Although Hurley's results were in accord with predictions made on the assumption that achievement scores reflect degree of motivation, they are also consistent with the notion that achievement scores reflect a general response tendency influencing performance in a wide variety of situations. Ss with high achievement scores may simply have the habit of being or at least appearing "eager." In other words, they may have learned certain work habits of an "effortful" or "active" sort. The distinction between drive-produced activity and learned activity has been nicely brought out by Miller (32), who points out that rats can be *trained* to be extremely active when hungry if pellets of food are widely scattered about a maze. Animals trained in this way would presumably be more active ("eager-appearing") than untrained animals, even though the latter might actually be hungrier. In like manner, human Ss appear to vary, as a result of learning, in the degree of obviousness of their desire to succeed. Under low motivational instructions Ss having relatively strong habits of this sort should tend to produce both more correct and incorrect responses. When the instructions themselves produce the "eagerness" response, the differences between the two groups might be expected to disappear. It is far from certain, then, that the term "achievement motive" is appropriate, at least when applied to the indices here considered; "achievement-oriented habit," though more awkward, may be more accurate.

IPIT hostility scores. Certain of

⁶ This was made possible by the generous cooperation of Dr. J. W. Atkinson, who furnished a prepublication copy of the manuscript.

⁷ The difference between the two correlations was significant at $p < .07$. The higher level of significance reported for this comparison in the original version of this paper was based upon a smaller N .

Hurley's results suggest that the personality (response-defined) variable operating in his study may actually have been hostility, rather than achievement-orientation. He found a highly significant negative correlation between the hostility scores on the IPIT and the number of both correct and erroneous anticipations in verbal learning, under low motivational instructions. This correlation disappeared under the highly motivating instructions, the difference between the two correlations being significant. Since the correlation between scores on the achievement and hostility indices of the IPIT was high ($r = -.71$), the relation of each score to performance was evaluated by using partial correlations. When the effect of the hostility score was eliminated, the partial correlations between achievement and performance were reduced to insignificance. However, when the effect of achievement imagery scores was partialled out, the negative correlation between hostility and number of correct and total (correct plus incorrect) anticipations under the low motivational conditions remained significant.

It is possible that high hostility, as measured by the IPIT, itself reflects a drive state. Since high *D* may, under certain conditions, be expected to produce poorer verbal performance (see below), such an assumption would not be inconsistent with Hurley's results. But this would imply that high achievement scores reflect low *D*. It seems simpler and more plausible, as Hurley himself concluded, to attribute the relations between the hostility scores and performance to differences in habit tendencies rather than to drive.

Taylor Anxiety Scale (A-scale) scores. A number of investigators in the Iowa Laboratory and elsewhere have been expressly concerned with

the experimental and theoretical problems involved in testing the supposition that a given response measure reflects drive. Using an anxiety questionnaire originally constructed and recently revised by Taylor (52, 53), these investigators reasoned as follows: If *Ss'* scores on this test reflect general drive level, and if variations in drive affect response strength in the manner suggested by Hull (23), then the performance of these *Ss* should differ in certain predictable ways in given experimental situations.

Since the studies were for the most part deliberately restricted to a consideration of nonassociative characteristics on the view that this procedure is most appropriate in a test for motivational properties, the assumption figuring most prominently in these predictions was that all habit tendencies elicited in a given situation are multiplied by the value of the total effective drive then operating. On this assumption, it can be shown that in learning situations in which correct responses are highly dominant, *Ss* with a relatively high drive level should perform better than those having a low drive level.⁸

If, however, erroneous response tendencies are stronger than the correct ones, the performance of *Ss* with high drive should be poorer than that of *Ss* with low drive. Furthermore,

⁸ Contrary to the supposition by Lazarus, Deese, and Osler (26), Taylor's prediction that anxious *Ss* would give more conditioned eyelid responses than nonanxious *Ss* was based upon this multiplicative assumption ($R = H \times D$), rather than upon the assumption that anxious *Ss* have the stronger associative tendency to make defensive reactions. In the case of simple eyelid conditioning, it is not possible, of course, to decide between these two interpretations of Taylor's original results (52). In other situations, however, the behavior predicted on the basis of these different hypotheses would presumably be quite different.

as Taylor and Spence (54) have pointed out, in complex situations, even though a correct response tendency might be slightly stronger than any alternative superthreshold tendencies, an increase in drive could nevertheless increase the frequency of erroneous responses. This does not mean that the presence of competing tendencies would always put Ss with high drive at a relative disadvantage, regardless of the strength of the correct tendencies, but, rather, that the relative strengths of the correct and incorrect tendencies are not the only factors to be considered. Thus, according to Hull's theory, an increase in drive (D) will increase the excitatory strength (E) of the strongest response more than that of the weaker ones, but will also increase the probability that more of the weaker responses will have superthreshold values. If the increase in D does not increase the E of the strongest tendency so much that its downward oscillation never brings it within the range of E 's for the weaker responses, the possibility of successful competition by weaker, but now superthreshold tendencies, may be increased. Clearly, there is no possibility of successful competition by the weaker tendencies if their strengths are subthreshold.

A number of deductions that could be made on the basis of the foregoing theoretical analysis have been tested in various situations, including verbal learning (33, 38, 48). The results have indicated that the quality of performance in complex learning situations is inversely related to Ss' degree of anxiety, as measured by the Taylor scale, and furthermore, that the advantage of the nonanxious over the anxious Ss is positively related to the probable number and strength of the competing responses elicited. At the moment of this writing it does not seem altogether cer-

tain whether conventional verbal learning tasks can be made simple enough to provide a consistent advantage for the anxious Ss. For instance, Montague's (33) anxious Ss were superior to the nonanxious on an easy list, consisting of items having high association values and a minimum of intralist similarity, but this simple effect was not significant; only the interaction between anxiety and difficulty of list was significant. However, more recent studies (see 48) indicate that it is possible to construct verbal learning situations that permit highly anxious Ss to reap the benefit to which their higher drive, if it indeed exists, should entitle them under appropriate conditions.

Theoretical implications of A-scale studies. It has not always been recognized that most of the Iowa studies of the relations between A-scale scores and various kinds of laboratory performance have been exclusively concerned with the test of the possibility that the A scale measures drive level (D). From the present point of view, the test of the supposition that the A scale or any other variable is a measure of D , or of motivation, generally, must be made in terms of predictions based on purely nonassociative assumptions, a procedure seldom if ever followed in the case of other response indices. It must be noted that this kind of test can be made only under certain conditions, namely, those in which the associative properties of the variable may be presumed to have little differential effect. Under such necessarily limited conditions, the general success of the predictions thus far may fairly be said to lend increased confidence in a drive interpretation of A-scale scores.

This conclusion does not imply that A-scale scores reflect nothing but D . On the contrary, if they reflect a

motivational state, then in the present view, they also undoubtedly reflect associative tendencies. Under many circumstances, therefore, *Ss* with relatively high scores should perform differently from those with low scores by virtue of differences in their habits. In such instances one could not expect to account for variations in the behavior of anxious and nonanxious individuals in terms of drive level alone.⁹

The theoretical studies concerned with the *A* scale were never intended, then, to investigate all of its possible functional relations. Nor, contrary to the belief of some, was the *A* scale constructed in order to predict as well as possible either eyelid conditioning, or verbal learning, or motor skill, or any other sort of performance. If this had been so, a very different procedure would undoubtedly have been followed. If one wished to use a self-rating questionnaire, one would administer a very large number of items, obtain the correlation of each with the behavior to be predicted, and then, after cross validation and a consideration of the

item intercorrelations, select the set that most efficiently predicted the criterial measure. A test constructed in this manner would unquestionably predict the specific kinds of behavior it was designed to measure better than the *A* scale could be expected to do.

It is certainly not surprising that various kinds of empirically selected measures excel the *A* scale as predictors of certain kinds of behavior. But since the *A* scale was not devised, nor ever considered by those associated with its construction, to be a measure of all of the variables influencing performance, some sort of limit is to be expected, under most conditions, in the amount of variance in performance that is associated with scores on the *A* scale. It is only in certain kinds of controlled situations that variations in performance are likely to be primarily due to variations in drive.

A misconception somewhat related to the notion that the *A* scale was intended to be a "best predictor" of some kind is that the "essential" validity of the Taylor *A* scale depends upon its correlation with other indices of anxiety, e.g., those of a clinical or psychiatric nature. There is some evidence, as a matter of fact, that scores on this test do relate to clinically diagnostic measures, including psychiatric ratings (17). Such relations, when they are found, are interesting and useful (though perhaps not so useful as one might wish, in view of the notorious unreliability of psychiatric judgments), but they most emphatically do not constitute a validation of the test in respect to the purpose it was originally designed to serve. As Bechtoldt (3) has recently observed, the question of the validity of the Taylor scale as a useful definition of general drive level is answered by the accuracy of the predic-

⁹ From the standpoint of these considerations, Child's (9) criticisms of the Iowa studies for failing to investigate the associative aspects of manifest anxiety do not seem to have been well-taken. Although the motivational status of anxiety could not be demonstrated by pointing to differences in the habits of anxious and nonanxious *Ss*, there is no doubt that such differences exist. On the other hand, despite considerable speculation and some empirical evidence (e.g., 3, 18, 19, 55, 56), it is not as yet entirely certain what these habitual differences may be, apart from the trivial observation that they consist, at least in part, in the kinds of verbal responses given on the test itself. For instance, despite some supporting evidence (20), it is doubtful, on either theoretical or empirical grounds, (49, 50) that a consideration of the specific nature of anxiety responses warrants the belief that anxious *Ss* are (habitually?) poorer than nonanxious *Ss* in discriminating between threatening and nonthreatening stimuli.

tion of relations between this scale and specified behavior variables, under conditions such that variation in the behavioral variables can be reasonably attributed to differential drive levels.

In view of this, the Taylor questionnaire might better have been termed a test of "excitability," or "level of responsiveness," or simply, in accordance with the present usage, an "A scale." Nevertheless, the items of the scale were selected by clinical psychologists in accordance with the specifications for manifest anxiety laid down by a descriptive psychiatrist of high repute (8). It does not seem inappropriate, consequently, to consider this test an index of manifest anxiety as well as D.

SUMMARY

It is the major thesis of the present paper that the question as to whether a given variable acts as a motive or reflects a motivational state can be answered only in the context of explicit assumptions concerning the particular ways in which motiva-

tional variables can be differentiated from other variables in their influence on behavior. In view of the current confusions between the concepts of motivation and habit, it would appear to be particularly desirable to distinguish between the associative and nonassociative properties of variables. Thus, experimentally manipulated or environmentally defined variables such as electric shock, hunger, fear, and failure may be classed as motivational by virtue of their non-associative (drive) properties; nevertheless, their effects upon verbal learning and performance in some situations are best understood in terms of their associative characteristics. With a very few exceptions, the influences of differentially selected or response-defined variables frequently classified as motivational also seem attributable, for the most part, to their associative properties; in the absence of convincing evidence that such variables have anything other than an associative influence, it is doubtful whether they qualify as motivational at all.

REFERENCES

1. AMSEL, A., & MALTZMAN, I. The effect upon generalized drive strength of emotionality as inferred from the level of consummatory response. *J. exp. Psychol.*, 1950, **40**, 563-569.
2. AMSEL, A., & ROUSSEL, JACQUELINE. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363-368.
3. BECHTOLDT, H. P. Response defined anxiety and MMPI variables. *Proc. Ia. Acad. Sci.*, 1953, **60**, 495-499.
4. BERGMANN, G. Psychoanalysis and experimental psychology: a review from the standpoint of scientific empiricism. *Mind*, 1943, **52**, 122-140.
5. BROWN, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1953. Pp. 1-21.
6. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465-480.
7. BROWN, J. S., KALISH, H. I., & FARBER, I. E. Conditioned fear as revealed by magnitude of startle to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317-328.
8. CAMERON, N. *The psychology of behavior disorders*. New York: Houghton Mifflin, 1947.
9. CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149-171.
10. CHILD, I. L., & WATERHOUSE, I. K. Frustration and the quality of performance: II. A theoretical statement. *Psychol. Rev.*, 1953, **60**, 127-139.
11. COMBS, A. W. The use of personal experience in Thematic Apperception Test story plots. *J. clin. Psychol.*, 1946, **2**, 357-368.

12. DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291-308.
13. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
14. ERON, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387-395.
15. FARBER, I. E. Anxiety as a drive state. In *The Nebraska symposium on motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1954.
16. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
17. GLENER, GOLDINE, & ULETT, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279-282.
18. GOODSTEIN, L. D. Interrelationships among several measures of anxiety and hostility. *J. consult. Psychol.*, 1954, **18**, 35-39.
19. HEINEMAN, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
20. HILGARD, E. R. Theories of human learning and problems of training. In *Symposium on psychology of learning basic to military training problems*. Panel on Training and Training Devices. Res. & Dev. Bd., 1953.
21. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229-234.
22. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
23. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
24. HURLEY, J. R. Verbal learning as a function of instructions and achievement motivation. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
25. LAMBERT, W. W., & SOLOMON, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269-279.
26. LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293-317.
27. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
28. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. (2nd Ed.) New York: Longmans, Green, 1952.
29. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1946, **56**, 244-251.
30. MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. (2nd Ed.) New York: Macmillan, 1950. Pp. 668-690.
31. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
32. MILLER, N. E. The role of motivation in learning. In *Symposium on psychology of learning basic to military training problems*. Panel on Training and Training Devices. Res. & Dev. Bd., 1953, 103-116.
33. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
34. MURPHY, D. B. Recognition thresholds and recall of nonsense syllables as a function of their frequency of occurrence, their after-effects, and of personality variables. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
35. NAGATY, M. O. The effect of reinforcement on closely following S-R connections: I. The effect of a backward conditioning procedure on the extinction of conditioned avoidance. *J. exp. Psychol.*, 1951, **42**, 239-246.
36. POSTMAN, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1953, pp. 59-108.
37. POSTMAN, L. Comments on papers by Professors Brown and Harlow. In *Current theory and research in motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1953. Pp. 55-58.
38. RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120-124.
39. ROE, ANNE. A psychologist examines 64 scientists. *Sci. Amer.*, 1952, **187**, 21-25.
40. SEARS, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33**, 229-274.
41. SHAFFER, L. F. *The psychology of adjust-*

- ment. New York: Houghton Mifflin, 1936.
42. SIEGEL, P. S., & BRANTLEY, J. J. The relationship of emotionality to the consummatory response of eating. *J. exp. Psychol.*, 1951, **42**, 304-306.
43. SIEGEL, P. S., & STUCKEY, HELEN S. The effect of emotionality on the water intake of the rat. *J. comp. physiol. Psychol.*, 1949, **42**, 12-16.
44. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
45. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47-68.
46. SPENCE, K. W. The postulates and methods of 'Behaviorism.' *Psychol. Rev.*, 1948, **55**, 67-78.
47. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.) *Comparative psychology*. (3rd Ed.) New York: Prentice Hall, 1951.
48. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Learning theory, personality theory, and clinical research: the Kentucky symposium*. New York: Wiley, 1954. Pp. 1-21.
49. SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399-403.
50. SPENCE, K. W., & FARBER, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127-134.
51. SPENCE, K. W., FARBER, I. E., & TAYLOR, ELAINE. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404-408.
52. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
53. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
54. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-66.
55. WESLEY, ELIZABETH L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129-134.
56. WESTROPE, MARTHA R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515-524.
57. WHITTAKER, EDNA M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732-733.

Received July 1, 1954

LEARNING DURING SLEEP?

CHARLES W. SIMON AND WILLIAM H. EMMONS

The RAND Corporation

Approximately 22 years of the average man's life is lost in sleeping. Economically minded persons and harassed students have long searched for some means to use this time to further advantage. For nearly a third of a century, now, there has been a growing interest in the possibility of trying to learn while one sleeps.

The science fiction writers were among the first to propose sleep-learning as an educational technique. In 1911, a magazine called *Modern Electrics* published a fiction story by Hugo Gernsback (4) in which the hero learned during sleep. The process was simple. Upon retiring, the student placed the material to be learned on a machine which played it automatically while he slept. In 1932, Aldous Huxley (9), in his *Brave New World*, scoffed at the possibility of improving the S's intellect while he slept, but described the world of the future as one which utilized the sleep period to train the lifelong attitudes of its populations. Other science fiction writers have continued to use the idea of sleep-training in their stories.

Popular news and picture magazines, citing experimental evidence, constantly reinforce the public's interest and misinformation about this osmotic form of education. From time to time, a national news service will release a story telling of some individual who works for a living by day and studies chemistry, operatic arias, or other lessons while sleeping at night. On "People Are Funny," a popular radio stunt show, an audience participant learned a few short sen-

tences and a girl's name while he was apparently asleep on the stage before the audience.

Following closely on the heels of public interest have come the entrepreneurs. Commercial firms have sprung up throughout the country selling recording devices which automatically turn on and off to permit the individual to learn while he sleeps. Impressive publications have been distributed by many of these firms propagandizing their products and extolling the validity and value of sleep-training (19, 21). On the whole, their claims of success for sleep-training have been extrapolations beyond the available supporting data. The majority of claims have been based on distorted facts, statements by unqualified authorities, and armchair hypotheses. Noncritical use of the information, anecdotal evidence, and the citing of inadequate research, much of which will be reviewed in the present paper, have made these commercial publications poor criteria for the support of the validity of the sleep-learning process.

In its constant search for new methods to speed training during national emergencies, the military, too, has tried to sleep-train its personnel. In both the first and second World War, exploratory efforts were made to teach service men Morse code^{1,2} during their sleep. A VA Hospital doctor reported the successful use of sleep-training as a supplement to waking-training in a mental health

¹ Comdr. R. R. Humes, USN. Personal communication. January 25, 1954.

² L. L. Thurstone. Personal communication. October 19, 1953.

program (20). Unfortunately, most of these incidents have been too inadequately controlled to judge their effectiveness during sleep, and in general their results have not been sufficiently striking to justify the extra energy and expense required to carry them out.

It is the scientist who has been slow in investigating these claims for successful sleep-learning. Although psychologists have been concerned with problems of learning from their formalized beginning, there has been relatively little interest in the psychological processes which take place during sleep and a rather skeptical—although untested—attitude toward the value of studying sleep-learning. One of Pavlov's students, Krasnogorski (11), did attempt without success to condition the salivary reflex in a young child during sleep. Recently, however, a resurgence of interest in the problem has occurred. In the present paper the available research on the sleep-learning phenomenon has been collected and subjected to a critical evaluation.

Ten such studies were found. Of these only three have been published (3, 12, 13). Five were academic theses at the master's or bachelor's level (1, 2, 5, 8, 23) of which one was read at an APA meeting (23). The remaining two^{2,3} were described in private communications from the authors. Although a number of other references were found, insufficient details were available for an adequate review or else they could not be classified as research under even the broadest definition.

The criticisms in this paper are not for the purpose of belittling these pioneer studies. These criticisms are believed necessary since these stud-

ies—many of which are unpublished—are being cited and their conclusions accepted both by the general public and some scientists. Because of the unusual methodological problems in this area of research, a review may help future workers to avoid similar mistakes.

Description of the Sleep-Learning Studies

The experiments reviewed in this paper are described briefly below. In some of the studies, Ss were eliminated for various reasons. Only the number of Ss actually completing the experiment is reported.

Thurstone,² in 1916, supplemented the waking training of 16 Navy men in a Morse code course with material presented during sleep. The criteria of sleep used in this study are unknown. The Ss finished the course three weeks earlier than had been expected. Thurstone concluded that the results "indicated some gain" for the sleep-trained group in their ability to send and receive Morse code.

LeShan (12), in 1942, tried to break the fingernail-biting habit of 20 boy campers, ages 8 to 14, by playing the phrase, "My fingernails taste terribly bitter," through a loudspeaker 300 times a night for 54 nights. The Ss were asked if they were awake before turning on the input; if the Ss appeared restless, the input was turned off. The nails of the Ss in the experimental group and an equal number of Ss in a control group were checked every two weeks for eight weeks for signs that the nail-biting had stopped. Since 40 per cent of the experimental group and none of the control group stopped biting their nails, LeShan felt this indicated "the possible therapeutic use of suggestion during sleep."

LeShan³ in 1943, taught a single S a different list of nonsense syllables each morning for 12 days. On the fifth and eighth nights, the list to be learned the next day was repeated 50 times while S slept. The criteria of sleep used in this study are unknown. Fifty per cent fewer trials were required to learn the two

² L. LeShan. Personal communication. October 27, 1953.

sleep-trained lists than to learn the ten nonsleep-trained lists. LeShan concluded that the sleep-training facilitated learning.

Elliott (2), in 1947, studied the effects of 30 repetitions of a list of 15 common three-letter words during sleep on the per cent saved in learning the list by the anticipation method the following day. Two groups of 20 male college students of equal learning ability spent one night adapting to the laboratory environment. The word list was repeated serially to one group of *Ss* while they slept. An EEG record was used as a criterion of sleep. No material was played when the *S* showed "clear alpha patterns" although the EEG was turned off when *E* believed *S* would stay asleep. On awakening the next morning, the two groups learned the sleep-trained list by the anticipation method and were asked if they awoke during the night. The group receiving the sleep-training showed a significantly greater percentage of savings than the control group ($p = .05$).⁴ No significant differences between groups were found on the basis of errors or absolute number of trials to learn. Elliot concluded that there is some retention of auditory material during sleep.

Hedges (5), in 1950, tried to improve the speech of three mentally retarded and aphasic children (ages 11, 7½, and 6) by sleep-learning. Short sentences, paragraphs, and simple consonant sounds were played to the children between 433 and 1501 times distributed over 7 to 13 nights. The machine was not turned off if an *S* awoke. The third child required fewer waking lessons to learn the sleep-trained consonant than it had required previously to learn a nonsleep-trained consonant. The babbling of the first child increased. The second child showed no effects. Hedges interpreted this as showing "the possibility of perhaps speeding" the training of the third *S*.

Fox and Robbins (3), in 1952, taught 30 college males and females a list of 25

pairs of English-Chinese vocabulary. The *Ss* were divided into three groups, matched on a pretest. Two vocabulary lists were prepared; in the first list, word pairs were matched as they were to be learned on the posttest (the facilitation list), while in the second, they were mismatched (the interference list). These lists were played 15 times to two of the *S* groups approximately three hours after they had retired to go to sleep. The third group—a control—heard only music while they slept. Although *Es* did not observe *Ss* during the input period, they disqualified all *Ss* who said they had awakened during the night. The facilitation group required fewer trials and the interference group required more trials to learn the posttraining list than did the control ($p < .001$). The *Es* concluded that learning could occur during sleep and could be detected by the savings method.

Leuba and Bateman (13), in 1952, believed that they had taught the words of three songs, 8 to 27 lines long, to a lightly sleeping *S*. Each song, unknown to *S* beforehand, was played to her five times a night for three successive nights as she slept. During the input period, *S* was not observed by *E*, although for three brief periods, *S's* husband observed no restlessness on the part of the *S*. The *S* claimed she did not awaken during the night. Given only the song title the next day, *S* was able to write the lyrics of two out of the three songs with no errors and of one with only minor errors. The *Es* believed the sleep-training was successful. No learning occurred in a later study during and following the use of sedatives.

Hoyt (8), in 1953, taught ten pairs of English-Chinese vocabulary to his *Ss*. Twenty *Ss* were matched on a pretest and given one night to acclimate to sleeping in the laboratory. On the experimental night, one group of eight *Ss* received twelve repetitions of a list paired in the same manner as that to be learned the next morning (facilitation list). A second group of eight *Ss* received a list of comparable words, but with the pairs mismatched (interference list). The remaining four *Ss* acted as a control group and received no sleep-training. During the input period, *Ss* were observed by *E*.

⁴ The p values representing the results of tests of significance are all reported as if they were based on a two-tailed t test, irrespective of what the author used. The p levels greater than .15 were not reported.

If they awoke or heard the stimulus material, they were to turn on a light. If this occurred, or if they stirred, the recording was turned off until *S* lay quiet for a suitable length of time. The next morning *Ss* were asked if they had awakened during the night. Those that had not were given the paired vocabulary list to learn to two correct anticipations. Statistically insignificant differences were found to favor greater savings for the interference group rather than the facilitation group in both the mean number of trials required to learn the list as well as the number of correct responses occurring before the criterion was reached. No comparisons were made with the control group. Hoyt felt that under the conditions of this study, learning during sleep could not be detected.

Stampfl (23), in 1953, had six college males learn lists of ten nonsense syllables while acting as their own controls. Different lists were repeated 4, 8, 16, and 32 times on different nights while *Ss* slept, and were learned to one correct anticipation on the morning following the repetition of a particular list. At other times, the same *Ss* were tested on other lists after no sleep-training. Before presenting the stimulus material, *Ss* were asked if they were awake. During the input period they were watched, and if movements occurred, the input was stopped. Since no significant differences were found for savings in either trials or errors between learning sleep-trained and non-sleep-trained lists, no comparisons were made among performances after different numbers of training repetitions. Stampfl felt that the sleep-learning hypothesis was uncertain and improbable.

Coyne (1), in 1953, carried out a series of exploratory studies on a variety of psychological problems using from four to six male college students as *Ss*. He believed the results to be generally favorable for sleep-learning. Unfortunately, a statistical error was discovered in Coyne's thesis after he had written his discussion and conclusions which led to a more favorable interpretation of the results than was justified. Although *E* could observe *S* as he slept, he primarily depended on *S* to press a buzzer as soon

as he awakened during the night. The problems Coyne studied are described below.

1. An interference list of adjectives was presented 25 times to the sleeping *S*. A similar list was learned to one perfect anticipation the next morning. A control group had no sleep-training. The sleep-trained group did poorer than the control group ($p = .07$).

2. When the first problem was repeated using *S* as his own control, more errors were made on the list which had received the interfering sleep-training ($p = .15$). No difference in savings (trials to learn the posttraining list) was observed.

3. Twenty pairs of numbers and words were presented 24 times during a single 45-minute period to the sleeping *Ss*. These same *Ss* also received another list for the same number of repetitions distributed over a four-hour period. On subsequent mornings, *Ss* were required to answer the appropriate word when given the number found on the list on which they were trained during the night, after that list had been mixed with ten additional new number-word pairs. A greater percentage of sleep-trained words were associated with the correct numbers than those words not trained during sleep ($p = .15$). Distributed sleep-learning resulted in fewer recall errors than did massed sleep-learning ($p = .05$).

4. While the *Ss* slept, one list of adjectives was repeated 100 times; on a different night, another list was repeated 25 times. On the morning following the sleep-training, the list was learned to one perfect repetition. No significant differences were found in performance between the two amounts of training.

5. The *Ss* were required to solve a number of problems by finding *E*'s solution of a concept composed of the correct combination of letters, colors, and their relative positions. The solution to one problem had been repeated 180 times while *S* slept the night before. Insignificant differences were found favoring the performance on the sleep-trained problem.

6. While *Ss* slept, one list of nonsense syllables was presented 30 times begin-

ning at two o'clock and, on a different night, another list was presented 30 times beginning at five o'clock. Since it was discovered that the two lists were of unequal difficulty, comparisons between performance at different times of presentation were not made.

7. During sleep, the contents of a particular picture were described to *Ss* 90 times. The following morning, a number of out-of-focus pictures were shown, including the one described during sleep. The degree of focusing required before *S* could identify the pictures was the measure of learning. No significant differences in the case of identifying sleep-trained and nonsleep-trained pictures were found either when *Ss* were asked to give unaided responses or when multiple-choice solutions were provided.

Rather than examine each of the ten sleep-learning studies independently, this paper has been organized to discuss them all on the basis of the following categories: experimental design, statistical considerations, methodological considerations, and sleep criteria.

EXPERIMENTAL DESIGN

The use of a control group or of using *S* as his own control is recognized as a necessary procedure in order to know whether a certain experimental effect is real or not. In a number of the sleep-learning studies, however, both of these techniques were conspicuously absent or inadequately handled.

Thurstone (see footnote 2) recognized the inadequacy of his *uncontrolled experiment* teaching Navy men Morse code and attempted to run a second study in order to compare the performance of one group which received sleep-training with another which did not. This study was discontinued, he reported, when it was discovered that ambitious Navy instructors of the control group had been giving extra daytime instruction in order not to be out-taught.

Hedges (5) used *no controls* with two of the three speech-defective children he attempted to sleep-train. Since improvement in speech is a maturational problem, even for retarded children, the need for a control was paramount. The increased babbling of the first child may have been independent of the sleep-training and due solely to an additional month or two of growth, although Hedges believed that it took place immediately after the introduction of the sleep-training. Since *S* also received waking-training on the same material, it is impossible to know to what extent one can attribute the increase in babbling to sleep-training. No apparent learning took place with the second *S*. Although Hedges' third *S* acted as his own control by learning to pronounce one consonant with sleep-training added to his waking-training and another consonant without the sleep-training, this was an inadequate control measure since the experimental design was such that the sleep-trained consonant followed the nonsleep-trained consonant. Any improvement in the latter could be attributed to practice as well as to maturation.

Nor was a control used in the work of Leuba and Bateman (13). In this study, *S* presumably had no previous knowledge of the songs played to her during sleep, yet was able to write the lyrics without further training on awakening. If this were the case, any learning which took place during sleep would be a significant improvement, although materials such as songs and poetry probably have some internal predictability. This form of experimental design represents an *implied control*. It is implied that *S* acted as her own control for had she been tested previously, no appropriate responses could have been made, nor could one suspect that maturation

tional factors were operating to produce positive results.

A number of *Es* actually used their *Ss* as their own control. In LeShan's (footnote 3) second study, he compared the number of trials an *S* required to learn a list with and without sleep-training over a period of days. By having the nontrained periods before and after the sleep-trained periods, the superiority of the sleep-trained lists could not be attributed simply to practice or maturation. On many of his studies, Coyne (1) gave sleep-training to his *Ss* one night and no sleep-training on the next, counterbalancing the order for these procedures between two groups of *Ss*. However, when the *S* acts as his own control, either the study material which is used under the varying conditions must be carefully equated, or additional counterbalancing must be introduced into the experimental design to correct for the inequalities. No *E* used this counterbalanced design; some made use of the published tables on which similarities and associability of the stimulus material had been previously calculated for the items. Several of Coyne's (1) studies were unanalyzable after he discovered the study material had not been equated.

Four of the experiments were designed to use *separate control groups*. In Elliott's (2), Fox and Robbins' (3), and Hoyt's (8) studies pretesting took place in order to equate the mean performance of the control and experimental groups. LeShan (12) used unmatched groups in his study with fingernail-biting children. The median age of his twenty experimental nailbiters was slightly less than ten years. His control was divided into two groups consisting of eight nailbiters with a median age of nine years, and twelve more nailbiters with a median age of twelve.

Some question might be raised concerning LeShan's failure to better equate the experimental and control groups on age, for there is reason to suspect a relation between nail biting and age. Wechsler (25) found a sharp rise in nail biting for boys around the age of twelve; if this is so, there would be a lower probability for the older control group to stop biting its nails than the younger experimental group, thus reducing the effectiveness of the older group as a control.

Two of the experimenters added a *second experimental group* to their design. Fox and Robbins (3) and Hoyt (8) used both a facilitation and an inhibition group in order to obtain a more sensitive indication of the value of sleep-training. Of all of the studies, only that of Fox and Robbins (3) provided the control group with a neutral stimulus—music—for the same amount of time as that in which the experimental group received the verbal test material. If there are any disturbances during sleep due to the stimulus and if these in turn affect recall, such a procedure is a wise one.

Although none of the *Es* were directly concerned with the problem, the use of an *additional control group* to compare results from sleep-training with the equivalent amount of waking-training would have been quite illuminating.

STATISTICAL CONSIDERATIONS

Only five of the *Es* (1, 2, 3, 8, 23) treated their data statistically to see if the sleep-training improved performance significantly. The remaining *Es* used clinical criteria to evaluate the effects of sleep-training.

Although Elliott's (2) results favored the performance of the sleep-trained group over the nonsleep-trained group, he failed to find a significant difference at the 5 per cent

level in the number of trials it took to relearn the training list. Elliott had equated his groups on a pretest, but did not attempt to match the individual Ss for the analysis. Since equating tends to decrease differences between means, failure to remove the variance due to Ss inflates the error variance and decreases the probability of getting significant differences. When the present authors (22) did an analysis of covariance with pre- and posttest scores from Elliott's data, the differences between the mean number of trials to learn a new list by sleep-trained and non-sleep-trained Ss were significant below the .05 probability level.

Coyne (1) used the *one-tailed t test* to evaluate his data. The wisdom of this treatment is questionable. In order to avoid abuses and controversy, exploratory work should be as cautious in its interpretations of results as it should be daring in attempting new ideas. Using the one-tailed *t test* does not allow for the possibility that differences might be in the direction opposite to that hypothesized (6). This is serious in any exploratory work; it is particularly dangerous in sleep-learning studies where one could seriously suspect that the intervening training during sleep might actually hinder normal waking recall.

Half of the Es (2, 3, 8, 12; footnote 2) used a reasonably large number of Ss as compared to the number used in most psychological studies; in the remaining cases, the number was smaller. When the number of Ss is small, one might be more critical of accepting the null hypothesis merely because the level of significance was not below the traditional 5 per cent level. For studies as exploratory as these, significance levels of 15 per cent could be arbitrarily considered encouraging. Since the expense in

time and money is relatively small, preliminary work in sleep-training should favor fewer Type II errors in order not to reject valuable experimental leads by accepting a false null hypothesis.

METHODOLOGICAL CONSIDERATIONS

A number of Es believed that differences in methodology might have been responsible for the divergent successes and failures found in the sleep-learning studies. These and other considerations are discussed below.

Subjects

The majority of the studies employed the traditional college student—male and female—as Ss. Thurstone used Navy men and LeShan used young boy campers. Hedges (5) bravely attempted to provide sleep-training as a supplement to the waking-training of children who had speech deficiencies and who were suspected of being mentally retarded. None of the Es attempted to study the effects of either age or sex on sleep-learning.

The selection of Ss may have an effect on whether successful sleep-training results are attained or not. Underwood, while reviewing Fox and Robbins' paper in the 1953 *Annual Review of Psychology*, commented that "such low variability [on performance scores] among Ss on the test list is rarely found in normal transfer experiments with such material, but this may again only reflect the presence of a highly select and homogeneous sample" (24, p. 48). Low within-group variability was certainly in part responsible for the high degree of significance of the differences between the sleep-trained groups and the control.

Whatever the variability of the group, it would appear wiser to

choose individuals who have shown the capacity to learn while awake. Perhaps the effects of sleep-training are so subtle that its benefits will be found only when it is applied to individuals with very high IQ's.

Number of Repetitions

The experiments can be divided into two groups on the basis of this variable—those who gave an exceptionally large number of repetitions of the material during sleep (5,12) and those that gave significantly fewer repetitions (1, 2, 3, 8, 13, 23; footnote 3). The large number of repetitions were actually spread over a number of nights and ranged from a total of 433 times over a period of eight nights to 16,200 times over a period of 54 nights. The smaller number of repetitions ranged from 8 to 180 times, with one study (13) playing the material five times per night for three successive nights. More repetitions were characteristic of the field as opposed to the laboratory studies. Both groups obtained both positive and negative results, although the tasks were varied sufficiently so that direct comparisons could not be made. LeShan's (12) study, in which 40 per cent of the experimental group stopped biting its nails, represented a successful example where a great deal of repetitious training seemed to have affected a semi-involuntary behavior.

Stampfl (23) and Coyne (1) found no differences when they tried to study the effects of different amounts of repetitions on sleep-learning. Actually since neither *E* used a non-sleep-trained group as a control, neither could conclude that any sleep-learning took place at all in this phase of their study. Coyne (1) suggested that in his study there may have been no greater differences in savings after one hour of repetitions than after four hours because the material was simple

enough to be learned in one hour and additional practice could not improve it. This is a reasonable hypothesis, as is its antithesis—that during sleep, learning is sufficiently slow so that little is learned even after four hours of repetitions. Sleep-learning, if it is to occur at all, may require that an extremely large number of stimulus repetitions take place. It will be important to evaluate sleep- versus waking-training from the standpoint of economy of both time and effort.

Presentation

The manner in which the study material is presented to the sleeping *S* has been considered by some *E*s as critical to the success or failure of the training. Two dimensions of this variable are the time of presentation and the order of presentation.

The problem of *presentation time* is an important one since during the sleeping period, time is related to some extent to the depth of sleep, which in turn may be related to trainability. Deeper sleep tends to be more prevalent in the early period of sleep, while lighter sleep tends to occur later (18). Of course, the levels vary considerably throughout a normal night's sleep.

Coyne (1), studying the effects of presenting the material at different times during the sleep period, failed to equate his lists beforehand and could draw no conclusions from his results. He recognized that presentation time might be inversely related to the amount retained. He suggested, however, that this was due to the recency of the presentation to the recall period. It is interesting to speculate whether or not the Jenkins-Dallenbach interference effect (10) occurs within the sleep period for materials presented during sleep.

The differences Coyne found between massed and distributed learn-

ing might also be accounted for on the basis of presentation time. Although he concluded that distributed sleep-learning was superior to massed sleep-learning, there is no way to determine whether this was so because of the spacing of the training or because some of the distributed inputs occurred during the period just before waking—often a light drowsy state—while the massed inputs occurred only during the deeper and possibly less receptive period.

The order in which the material was presented during sleep may also affect the results of sleep-training. In a waking state, serial learning has been shown to be easier than learning material presented in a varying order (7). Thus, if any learning takes place in sleep, a serial presentation would more probably increase any positive effects which might occur. Also, if the sleep state is one in which no mental organization takes place, this would favor the learning of only the more organized serial presentation.

Hoyt (8) and Stampfl (23) varied the order in which their material was presented and failed to find that any sleep-learning took place. Fox (3), LeShan (12), Elliott (2), and Leuba (13) believed they found evidence of successful sleep-learning. Thurstone probably varied the order of his material, but his Ss were practiced over a period of months so that the effects of presenting the material in a varying or an unvarying order may have been minimized.

The varying presentation order in Hoyt's study was methodologically different in one major respect from that used in Stampfl's study. Since the former study used paired-associate material, varying the order in which the pairs were presented would have less effect than it had in the latter study where the order of words in a list was varied during sleep-training,

even though they had to be recalled serially during the waking period. Any positive effects of sleep-training in the latter case may have been neutralized by the negative effects built up through nonserial learning.

Training Problems, Materials, and Mode of Input

The types of psychological problems studied by the majority of sleep-training investigators have been quite limited. With the exception of Coyne's (1) work on concept formation and perceptual sets and LeShan's (12) and Hedges' (5) therapeutic studies, the remainder of the research has been involved with training problems which require the memorization of word lists. It is unlikely that all types of problems are suitable for sleep-training, although exploratory studies should examine many rather than a few possibilities.

Of the material used in the studies requiring verbatim recall, the degree of meaningfulness ranged from lists of nonsense syllables through short words to foreign language vocabulary. We know that the more meaningful the material, the easier it is to learn in the waking state (7). Stampfl (23) suggested that this might explain why his results with nonsense syllables were poorer than Elliott's (2) who used a list of adjectives, and why Fox (3), using a Chinese-English vocabulary, got even more striking results. This does not seem to be a critical variable, however, for Hoyt (8), using the same Chinese-English vocabulary as Fox (3), got negative results, while LeShan (footnote 3) and Coyne (1), using nonsense syllables, got positive results.

All of the Es used an auditory input. This is certainly the most obvious technique and would appear to be the most economical; however, other sensory channels need not be

ignored. It may be necessary to flash lights on closed eyelids or to apply tactual stimuli to the fingertips in code in order to "reach" the sleeping S.

Techniques and Measures of Retention

The techniques used to measure the retention of training material can have considerable influence on the amount of material recalled. However, in the present studies, there did not appear to be a pattern of successes or failures consistent with the technique used.

A number of *Es* did not require the verbatim recall of the verbal material presented during sleep. Instead, they evaluated retention on *S*'s ability to use the material in posttraining tasks or on an observed change in *S*'s behavior after the training. LeShan (12) examined his *Ss*' fingernails every two weeks to see if the children had stopped biting their nails. This technique might have been slightly more objective had the examiner not known which children were and were not receiving the sleep-training. Hedges' (5) clinical evaluations of his children's speech improvement required even more subjective judgments. Because of the very complexity of this measure, Hedges wondered whether it was sensitive enough to detect improvement. Coyne (1), in several studies, had his *Ss* describe an out-of-focus picture, the identity of which had been given to them during sleep, and to determine *E*'s solution of concepts composed of the correct combination of certain stimulus variables, the answer to which had also been provided during sleep. Thurstone's *Ss* were evaluated on their ability to send and receive Morse code. Both Coyne's and Thurstone's results could be quantified.

Leuba (13) gave his *Ss* the titles of songs played during sleep and re-

quired them to recall unaided the lyrics. Positive results were claimed for sleep-training in this study.

The savings method has been used by a majority of *Es* (1, 3, 8, 23; footnote 3), for they believed it to be a more sensitive measure of retention. Stampfl (23) believed that although material could not be consciously recalled at any moment, its presentation during sleep may have still modified the nervous system sufficiently to make learning easier and a savings effected. However, use of the savings technique may confuse the measure of retention with a measure of the ability to learn since both are confounded within the same performance score. None of the *Es* using the savings measure compared the time to relearn a list with the time it took the same *S* to learn an equivalent list presented after sleep. Comparisons with the original list presented before sleep are not sufficient. The positive transfer which occurs from the pretest to the posttest—the phenomenon of "learning how to learn"—may account for some of the apparent savings which several researchers believed they found. However, Stampfl (23) gave three practice lists to be learned before the experiment in order to bring *Ss*' learning curve closer to its asymptote and thus reduce the "learning to learn" effect.

A still more sensitive technique for measuring retention is that of *recognition*; at least in Luh's (16) classical study this was so after a two-day interval between training and the post-training test. Coyne (1) used this technique in the form of a multiple-choice test for one of his exploratory studies on perceptual set, but still failed to get positive results. Hoyt (8) told of an exploratory study in which he presented a single number-word pair to a sleeping *S* who was under constant observation for a total

of 11 hours on three successive nights. On awakening after the third night, *S* was given the number and asked to pick the correct word from a group of ten. He was unable to do this correctly. A major difficulty with the recognition method is that it must be corrected for chance guessing. With words, however, this correction for chance is difficult since all words do not have an equal probability of being recognized (17).

Variations in the score used to measure retention have failed to consistently differentiate success and failure in sleep-learning. Many of the *Es* (1, 2, 8, 23; footnotes 2, 3) used the *number of trials* to learn posttraining material to one perfect repetition as a measure of retention. Hoyt (8) required that the lists be learned to two perfect anticipations; this may tend to make his results slightly more reliable. Some *Es* (1, 2, 8, 23) also used as measures of retention the *number of errors* or correct responses made on the first trial or later trials, or on cumulative trials. Where the *Ss* performance was used to measure retention, the measure was characteristic of the task, e.g., Coyne used the extent of focus of the projector as the measure of performance in his perceptual set study. The measure that is used may determine in part the results which are found. For example, Coyne (1) noted that the measure by trials often favored the opposite results than those favored by the measure by errors. Error measures generally lead to less conclusive results than those measured by trials. This situation does occur in waking-research and need not be contradictory, although it certainly affects the conclusions drawn as well as the practical applications of the results.

Before final conclusions can be drawn concerning the feasibility of

sleep-learning, more recall techniques should be tried. Simon and Emmons (22) discuss this while considering the possibility of secondary cerebral storage mechanisms for material introduced during sleep.⁵

SLEEP CRITERIA

Perhaps the most damaging criticisms of the sleep-learning studies to date have been the inadequate control of sleep and the criteria used for defining sleep. So elusive is the process of sleep from the psychological standpoint and so little is known about the actual physiological mechanisms which cause sleep that the problem of determining whether *S* is asleep is to a great extent a semantic one. Although both direct observation and *S*'s subjective report are reasonably reliable for deciding if *S* has been asleep over a block of time, neither can be considered sufficient to know if *S* is asleep at any moment in time. Therefore, the care with which *E* determines the sleep condition of his *S* at the time of the input is highly important and determines the degree of confidence which one may place in the conclusion that learning during real sleep is or is not possible. Some *Es* (1, 2, 3) eliminated

⁵ Dr. Bernard Fox, in a private communication (26 June 1954), described his attempt to use hypnosis as a means of facilitating the recall of sleep-trained material. A 55-year-old man was brought to a point where deep hypnosis was possible. In his normal sleep he was presented with a list of ten words repeated for a half-hour. During this period, *S* was observed and the input turned off whenever he moved. The next morning, he was unable to recall any words unaided. When he was hypnotized, he made only one association which *might* have been related to one of the words. Still under hypnosis, he recognized fewer words than would be expected by chance when the ten were interspersed randomly in a group of thirty. Neither reading the words aloud nor making him choose by a forced-choice method produced any more positive results.

S from the experiment if he awoke or said he heard the stimulus. Other *Es* (12, 23) shut off the stimulus until *S* went back to sleep.

In four of the studies (2, 3, 8, 13), *Es* asked *Ss* after the training if they had awakened. This is an unsatisfactory criterion of sleep for experimental studies since it is a common experience to awaken during the night, perform a number of rational acts before retiring again and remember nothing the next day (8, 18). Hoyt stated that one *S* was observed to awaken during the night, leave the laboratory to go to the bathroom, and later return to bed. The next morning, *S* remembered nothing about it. In two of these studies (3; 13) *Es* were not even present to observe *Ss* as an added check.⁶

In one experiment (3) the time of night was considered a partial criterion of sleep; training was applied during the period when studies have shown sleep to be at its deepest. However, published sleep curves are based on statistical averages and cannot be used to predict accurately the sleep characteristics of the single individual. Individual sleep curves vary considerably from time to time during the night (15).

Two of the *Es* (12, 23), before presenting the stimulus material, asked *Ss* if they were asleep. Since LeShan's (12) *Ss* were adolescent boys in camp, naive as to the purpose of such questioning, one may wonder if asking them if they were asleep would be a sufficient check. It has

been observed that even willing *Ss* often fail to respond to such questions, although they admit hearing them later. Another major difficulty with this technique is that *S* may actually have been asleep when the test question was asked, but later would awaken sufficiently to hear the subsequent input. In these studies, no systematic check was made to be sure that *Ss* did not awaken after the initial test, except to note if they moved.

Coyne (1) and Hoyt (8) had their *Ss* press a button whenever they awoke in order to stop the input of training material. This method alone is not sufficient since it has been found that often the desire as well as the ability to do this is not present although *Ss* are able consciously to hear and understand the material. Since much of the material used in these studies was short and specific, even a few seconds of waking presentation would have been sufficient to allow *Ss* to hear and learn much of the stimulus while awake.

Stampfl (23), LeShan (12), and Hoyt (8), in addition to asking their *Ss* if they were or had been awake, also noted whenever they moved, and turned off the machine at that time. This criterion of sleep is not completely satisfactory since EEG records indicate that sleep may lighten to almost a wakening state before movement occurs, as well as afterwards, or when there has been no movement at all (15). Hoyt (8) found that two of his *Ss* said they had awakened but showed no observable movements, while three subjects said they had not awakened but showed from two to six movements each. This suggests that the criteria of asking and noting movements did not correlate very well. In Hedges' (5) study, the parents occasionally observed the children to see whether

⁶ Fox stated "if any *Ss* in the group of 30 actually did wake up without reporting it, then the number was probably distributed about equally throughout the groups. Moreover, since all *Ss* with one exception show effects in the expected directions, it is clear that the results cannot be explained on the bases of a few *Ss* in each group who awakened but failed to report doing so" (3, p. 78). It may be on the basis of many.

or not they awoke during the sleep-training; two children awoke. In the first case the number of training repetitions were not counted for the time *S* was awake. However, without a waking control, this correction is meaningless. The other *S* awoke while the recording was playing and began drowsily to follow the instructions given on the record. Hedges felt this aided the progress observed in that case.

Elliott (2) used the *electroencephalogram* to determine the *S*'s sleep level. Independent workers using a variety of techniques have established a significant correlation between the brain wave patterns and the depth of sleep (14). Because the writers believe that properly evaluated electroencephalograms represent the most objective, continuous, and practical on-the-spot indicator of sleep depth available, Elliott's (2) positive results—often quoted by the commercial firms in support of their claims for sleep-learning—are quite noteworthy. An examination of his sleep criterion revealed one major flaw, i.e., Elliott did not keep the EEG running during the entire training period. Therefore, no continuous check was made of the *S*'s sleep level while the stimulus was being played. The EEG was turned on at the beginning of the training and kept on until *E* believed that sleep would remain deep enough; then he would shut it off for the remainder of the half-hour stimulus period. With Elliott's data, a correlation of $-.39$ was found between the amount learned and the amount of monitoring. This was not significant at the 5 per cent level but in a direction which suggested that more savings occurred with less monitoring. The writers (22) also compared the amount of savings made by a group

of Elliott's five *S*s who were monitored nearly all of the training time with the savings of the remaining fifteen *S*s who were monitored on an average slightly more than a fourth of the time, and never more than 54 per cent of the time. For the first group, the average savings in trials to learn was 27.5 per cent; for the second group, the average savings was 45.0 per cent. The difference of 17.5 per cent was highly significant below the 2 per cent level of significance and we can conclude that when a thorough check of the sleep condition was made with the EEG during training, considerably less savings took place.

That any savings occurred for the groups monitored nearly 100 per cent of the time might be accounted for in two ways without assuming that learning actually occurred during sleep. First, the apparent savings may be simply an effect of the "learning how to learn" phenomenon discussed earlier in the section on "Methodological Considerations." Second, Elliott stated that whenever an alpha pattern was "clearly" present he concluded *S* was awake and turned off the input machine. In so doing, he quite likely permitted many of his *S*s to listen while they were already awake since movement, tenseness, or the opening of the eyes may sometimes block a clearly established alpha.

The expense of the EEG equipment negates its widespread use. This does not mean that without it, however, adequate sleep-learning studies cannot be done. The use of a combination of the criteria and continuous monitoring may be sufficient to insure some rough but adequate control of *S*'s sleeping condition. It is interesting to note that Hoyt (8) and Stampfl (23), finding negative results, were the two *E*s using the

greatest number of multiple and continuous criteria.

The continuity of monitoring cannot be overemphasized; for in spite of wishful thinking to the contrary, Ss do awaken while the input is being presented. Two out of three of Hedges' (5) Ss awoke. As many as half of Coyne's (1) Ss awoke in some of his studies. Elliott (2) reported that six of his Ss awoke. Only two out of sixteen of Hoyt's (8) Ss in the experimental group failed to awaken during an input period. LeShan (12) also stated his Ss awoke, but did not specify how many. One-fourth of Fox's (3) Ss awoke, and were eliminated from the study.

The conditions under which Ss sleep can influence to some extent the soundness of their sleep. Elliott (2) and Hoyt (8) required Ss to sleep in the laboratory one night previous to the experiment. Any sleeplessness occurring on this first night probably induced a deeper sleep on the second. Hedges (5), Fox (3), and Leuba (13) used Ss while they slept in their own homes, while Thurstone (footnote 2), LeShan (12), and Stampfl (23) used Ss who had been sleeping in familiar "homes away from home" for some time.

Although most of this discussion has centered around the problem, Was the sleep state deep enough?, Stampfl (23) and Leuba (13) suggested that there is an intermediate point between waking and deep sleep that is optimal for sleep-learning. Some negative results, they believed, may have been obtained because S was too *deeply* asleep. Whether this

is true or not is an experimental question for future Es to answer.

SUMMARY AND CONCLUSIONS

Ten sleep-learning studies were reviewed. Many of these have been cited uncritically by commercial firms or in popular magazine and news articles as evidence in support of the feasibility of learning during sleep. A critical analysis was made of their experimental design, statistics, methodology, and criteria of sleep. All of the studies had weaknesses in one or more of these areas.

It is highly speculative whether or not the studies reviewed in this paper have presented any acceptable evidence that learning during *sleep* is possible. The inadequate control of a number of experimental variables makes the validity of the conclusions drawn by many of the Es unwarranted. The conditions under which the results were found tend more to support the contention that some learning takes place in a special kind of waking state wherein Ss apparently do not remember later on if they had been awake. This may be of great practical importance from the standpoint of economy in study time, but it cannot be construed as *sleep-learning*. More carefully controlled experiments in the future may provide us with a clearer answer to the question, "Can one learn during sleep?" as well as to provide comparative data between waking and resting learning from the standpoint of economy of time and effort. The problem is partially confounded by an inadequate definition of sleep.

REFERENCES

1. COYNE, M. L. Some problems and parameters of sleep learning. Unpublished honors thesis, Wesleyan Coll., Conn., 1953.
2. ELLIOTT, C. R. An experimental study of the retention of auditory material presented during sleep. Unpublished master's thesis, Univer. of N. Carolina, 1947.
3. FOX, B. H., & ROBBINS, J. S. The retention of auditory material presented during sleep. Unpublished master's thesis, Univer. of N. Carolina, 1947.

- tion of material presented during sleep. *J. exp. Psychol.*, 1952, **43**, 75-79.
4. GERNsBACK, H. Ralph 124C 41+. *Modern Electrics*, 1911, **4**, 165-168.
 5. HEDGES, T. A. The effect of auditory stimuli presented during the sleep of children with delayed speech. Unpublished master's thesis, Univer. of Wichita, 1950.
 6. HICK, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, **59**, 316-318.
 7. HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517-567.
 8. HOYT, W. G. The effect on learning of auditory material presented during sleep. Unpublished master's thesis, George Washington Univer., 1953.
 9. HUXLEY, A. L. *Brave new world*. New York: Doubleday Doran, 1932.
 10. JENKINS, J. G., & DALLENBACH, K. M. Obliviscence during sleep. *Amer. J. Psychol.*, 1924, **35**, 605-612.
 11. KRASNOGORSKI, N. I. Bedingte und Unbedingte im Kindesalter und ihre Bedeutung für die Klinik. *Ergeb. d. Inner Medis. und Kinderhk.*, 1931, **39**, 613-730.
 12. LESHAN, L. The breaking of habit by suggestion during sleep. *J. abnorm. soc. Psychol.*, 1942, **37**, 406-408.
 13. LEUBA, C., & BATEMAN, D. Learning during sleep. *Amer. J. Psychol.*, 1952, **65**, 301-302.
 14. LINDSLEY, D. B. Electroencephalography. In J. McV. Hunt (Ed.), *Personality and behavior disorders*. New York: Ronald, 1944, Pp. 1033-1103.
 15. LOOMIS, A. L., HARVEY, E. N., & HOBART, G. A. Cerebral states during sleep, as studied by human brain potentials. *J. exp. Psychol.*, 1937, **21**, 127-144.
 16. LUH, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 3.
 17. MILLER, G. A. Speech and language. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 789-810.
 18. MULLIN, F. J., KLEITMAN, N., & COOPERMAN, N. R. Studies on the physiology of sleep changes in irritability to auditory stimuli during sleep. *J. exp. Psychol.*, 1937, **21**, 88-96.
 19. POWERS, M. *Mental power through sleep-suggestion*. Los Angeles: Wilshire Book Co., 1949.
 20. SCHMIDHOFER, E. Mechanical group therapy. *Science*, 1952, **115**, 120-123.
 21. SHEROVER, M. Learning during sleep: is it possible? New York: Modernophone, Inc., 1954.
 22. SIMON, C. W., & EMMONS, W. H. Considerations for research in a sleep-learning program. Santa Monica, California: Rand Corp. Paper No. 565, 1954.
 23. STAMPEL, T. The effect of frequency of repetition on the retention of auditory material presented during sleep. Unpublished master's thesis, Loyola Univer., Chicago, 1953.
 24. UNDERWOOD, B. J. Learning. *Annu. Rev. Psychol.*, 1953, **4**, 31-57.
 25. WECHSLER, D. The incidence and significance of fingernail biting in children. *Psychoanalytic Rev.*, 1931, **18**, 201-209.

Received July 15, 1954

TEACHING IN THE IVORY TOWER, WITH RARELY A STEP OUTSIDE

S. L. PRESSEY

Ohio State University

In recent years, psychologists have literally gone all over the world, and addressed themselves to a great variety of problems. In so doing, they have tried a variety of training and instructional methods, often on the job or in the field, and have evaluated those methods realistically, as with reference to outcomes in production records or combat data. One might hope that all this would have invigorating influences upon methods in the teaching of psychology, and appraisals of them.

Instead, an extensive and careful recent review¹ of investigations of the teaching of the subject since 1942 shows the largest single group dealing with phases of the old comparison of lecture versus discussion in the elementary course, as appraised most commonly by final examinations. And the reviewers seem to excuse the minimal imagination as to methods and minimal results by implying a minimal "degree and nature of change that one can produce in a one- or two-semester course in psychology" (p. 52). But why are not these limitations of time a challenge to use varied and vigorous methods, chosen on the basis of incisive evaluation? Yet the review seems not to mention any use of laboratory methods, nor any projects, field trips, or field experience. And the great majority of the evaluations seem to have been in terms of paper-and-pencil tests, usually given in the evaluated course (not later or elsewhere). Only talk-

talk in a college classroom as a means of bringing about learning, and only markings there on a piece of paper to evidence whether significant learning has occurred!

But many other methods are possible, even in a big elementary course—as the writer knows from over 25 years experience supervising such a course, enrolling some thousand students a year (mostly freshmen), taught mostly by graduate students. There can be informal and social laboratory tasks, ranging from genuine research projects which the students understand as such and are later reported in print, or the making of simple equipment such as charts for the course, to informal half-humorous data gathering regarding the "social psychology" of freshmen. There can be field trips, as to an institution, or field work as helping in the survey of a public school. And evaluations can include careful checking that the laboratory research is well done and the charts well made, notations that the school principal spoke favorably of the survey assistants and that after an institution visit, a girl asked intelligent questions about career opportunities in work there, or an anecdotal record that on a field trip an initial isolate had become an accepted member of a group.

The Wolfe committee doubted the legitimacy of many such doings as either too elementarily useful or too special: one should "teach psychology"; and only certified clinicians or counselors could properly deal with many of these problems. But beginning courses serving mostly stu-

¹ BIRNEY, R., & MCKEACHIE, W. The teaching of psychology: a survey of research since 1942. *Psychol. Bull.*, 1955, **52**, 51-68.

dents who will not go further in the subject surely should be realistic about them and their needs—can thus best serve them *and* give them some functioning understanding of the subject. And bringing it about that an isolated student gets acquainted need require little more of a good instructor than being a good host; even aid in early consideration of a vocational choice sometimes requires little more than mature good sense, with referral.

The reviewers report the extraordinary finding that sophomores get better marks in a first course in psychology than do freshmen, and seem to agree with the inference that the first course should therefore not be given in the first year. But if the first course were not taken till graduate school, grades would be still better! Surely not where students do best is the place for a course, but where they need it the most.

The rather unhappy reviewers conclude with the statement that the

most serious need, in research in teaching, is to get it "related in some way to general psychological theory" (p. 64).² Climb higher in the ivory tower! Why not instead come down and open the door, watch what students do outside class, see what psychologists are doing in the wide world, and even venture off campus occasionally, to mix with the folks on Main Street.

Received March 24, 1955

² There could well be more consideration of a broad theory of higher education. More than any other subject field, psychology *should* have appreciation of the great variety of gains—in abilities, understandings, personality traits—possible during the college years. Psychologists have a great variety of contributions to make, in furtherance of those gains. Some are highly special, as in counseling or therapy. But the basic contention of this commentary is that there remains a great range of possible contributions, most best attempted by broadened concepts of and methods in teaching—and that the excellent studies summarized in the review cover that range only from A perhaps through C. The plea is for a wider range of experiment.

ADDITIONS TO PSYCHOLOGICAL NECROLOGY: JAPAN, 1928-1952

KOJI SATO

Kyoto University

In the former Psychological Necrology (1903-1927) by E. G. Boring (2) two Japanese psychologists were listed, namely, Y. Motora and T. Nakashima, but in the new one by Bennett and Boring (1) none were included. Perhaps this is because communication between American and Japanese psychology was interrupted during World War II. However, Japanese psychology has developed a great deal during this period. Among the deceased psychologists in Japan there are a number who are worthy of being included in Boring's new list. From a total of 45 deceased psychologists, I selected 25 names and asked the leading psychologists in Japan, Tatsuro Yatabe (Kyoto Univer.), Toru Watanabe (Nihon Univer.), Matsusaburo Yokoyama (Keio Univer.), Sadaji Takagi (Tokyo Woman's Christian College), Megumi Imada (Kanseigakuin Univer.), Yoshikazu Ohwaki (Tohoku Univer.), and Torao Obonai (Tokyo Univer. of Education), to check them against the level of Boring's list. The following 13 were selected:

- Fukutomi, Ichiro, Kenkoku Univer., Manchuria, b. 8 June 1891, d. 27 Feb. 1946.
- Hayami, Hiroshi, Keijo Univer.,

- Korea, b. 23 Oct. 1876, d. 27 June 1943.
- Iwai, Katsujiro, Kyoto Univer., b. 17 Feb. 1886, d. 2 Nov. 1937.
- Kakise, Hikoza, Ministry of Education, Tokyo, b. 13 Sept. 1874, d. 23 April, 1944.
- Kuroda, Ryo, Keijo Univer. and Gifu City, b. 30 Jan. 1890, d. 5 Jan. 1947.
- Kubo, Yoshihide, Hiroshima Bunrika Univer., b. 22 Apr. 1883, d. 12 Nov. 1942.
- Masuda, Koresbige, Tokyo Univer., b. 29 Dec. 1883, d. 7 Aug. 1933.
- Matsumoto, Matataro, National Academy of Science, Tokyo, b. 15 Sept. 1865, d. 24 Dec. 1943.
- Morita, Shoma, Jikeikai Medical School, Tokyo, b. 18 Jan. 1874, d. 12 Apr. 1938.
- Nakamura, Katsumi, Tokyo Univer. of Education, b. 7 Feb. 1901, d. 6 Apr. 1952.
- Onoshima, Usao, Ministry of Education, Tokyo, b. 16 Mar. 1894, d. 24 Jan. 1941.
- Rikimaru, Jien, Taihoku Univer., Formosa, b. 3 Jan. 1890, d. 4 Dec. 1945.
- Takashima, Heizaburo, Toyo Univer., Tokyo, b. 1 Oct. 1865, d. 15 Feb. 1946.

REFERENCES

1. BENNETT, SUZANNE, & BORING, E. G. Psychological necrology (1928-1952). *Psychol. Bull.*, 1954, 51, 75-81.¹
2. BORING, E. G. Psychological necrology (1903-1927). *Psychol. Bull.*, 1928, 25, 302-305, 621-625.

Received March 15, 1955.

¹ The authors of this article welcome the opportunity to correct one gross error in their

list. Narziss Ach was born on 29 October 1871 (not in 1848).

SPECIAL REVIEW
HANDBOOK OF SOCIAL PSYCHOLOGY¹

DANIEL KATZ
University of Michigan

A handbook in any field of knowledge is an ambitious undertaking. When that field is growing so rapidly that changes of major significance take place between the writing of a manuscript and its publication and when the field ramifies widely with ill-defined boundaries, the task is even more formidable. In this perspective the two volumes edited by Lindzey are a magnificent achievement. They replace all existing texts and sourcebooks or any combination of text and sourcebook. They present clear and incisive accounts of the major theoretical positions in social psychology and they give good coverage to the thinking and research on all the traditional topics in the field. Many of these topics are old chapter headings in social psychology (humor and laughter, facial expression, mass phenomena) which more recent texts have omitted. And there is an excellent section of nine chapters on research methods. Moreover, many of the chapters are more than reviews of the literature in their insightful integration or in their suggestions for future research. Though social psychology has been characterized by sectarian trends and an impatience with an historical understanding of its problems, the *Handbook of Social Psychology* should help to counteract these forces. Some of the chapters are still too selective but the over-all pattern is one of real progress. High

editorial standards have been maintained and the writing is not only clear, but for a handbook, unusually interesting. Almost any adequate organization of chapters in this field would have problems in finding mutually exclusive areas but the amount of overlapping in the literature reported by the various authors is not great.

The opening chapter by Gordon Allport presenting the historical background of modern social psychology is one of the finest in the two volumes. It is a convincing argument for the importance of the historical approach. It shows the similarity and, sometimes, the virtual identity of contemporary concepts and problems with the ancient history of our discipline and demonstrates how easily some of the blind alleys in our development might have been avoided with more consideration of the thinking and work already available. If its wise precepts were to be followed we might temporarily solve our overcrowded journal situation by seeking for significant research problems rather than following the latest fad in labels and techniques. This chapter is rightly titled "Historical Background" in its concentration on the similarities between current ideas and their early counterparts. What would also be desirable would be a companion chapter on historical development which with the same bold strokes would show how we have moved from the early stages to our present situation. This, however, is not the conventional task of

¹ G. Lindzey (Ed.) *Handbook of social Psychology*. Cambridge: Addison-Wesley, 1954. Vol. I (pp. 588+x), Vol. II (pp. 601-1226+x), \$8.50 one volume. \$15.00 per set.

a handbook. The reviewer would make one minor criticism reflecting his own bias. There is too little material on the objective approach represented by such writers as F. Oppenheimer, E. Durkheim, K. Marx, and E. B. Holt.

Of the five chapters devoted to contemporary systematic positions, those on field theory and on Freudian theory are superior presentations in their clarity, organization, and relevance to social psychological problems. M. Deutsch does a remarkable job of exposition in condensing Lewinian metatheory, the formal concepts of field theory, and the research contributions of this school within the confines of a single chapter. His judgment that the experimental contributions of the movement stem from Lewinian metatheory rather than from Lewin's brilliant insight into empirical problems is probably sound. This does not mean, however, that the significant research activities of the school are the outcome of a formal set of theoretical concepts. The lasting contribution of Lewin thus may be a way of thinking about problems rather than a theoretical system. This is suggested, too, by the fact that Lewinian students have gone off in many different directions with some identification with a point of view and less adherence to the development of a common set of formal concepts. An interesting discussion in the Deutsch chapter centers about the problem of psychological ecology—namely, how field theory can take account of objective environmental forces, for in field theory these forces must be transformed into the psychological environment. The author accepts the concept of Wright and Barker of the *behavior setting* as "any part of the nonpsychological milieu that

is generally perceived by the people of a community as appropriate for particular kinds of behavior." Such behavior settings are coercive because of the shared frames of reference or collective intentions concerning appropriate behavior. This is one way of defining social norms and of accounting for the consistency of the resulting behavior. But it does not help on the problem of ecology. It still must be established that each individual subject has the same expectancies as his fellow and that he knows that his fellows have similar frames.

The Hall-Lindzey treatment of Freudian theory and its influence on social psychology is again an excellent summary and interpretation. If anything, it is on the conservative side in evaluating Freudian influence. Even where social psychologists have not accepted Freudian theory as a system they have utilized extensively Freudian concepts and Freudian mechanisms. Though their use of projection, displacement, and identification have not always accorded with Freudian theory, they have come to fall back upon Freudian dynamics for purposes of explanation. The early impact of Freudian theory, as in F. H. Allport's *Social Psychology* and E. D. Martin's *Crowd Behavior*, is also minimized in this chapter. It might be added that the tremendous vitality of Freudian theory is due in part to the fact that its mechanisms have some basis in the actual observation of human behavior.

In reporting on major trends in cognitive theory, M. Scheerer skillfully analyzes some of the basic issues about which stimulus response and organizational theorists have been in disagreement. The similarities of the concepts employed by cognitive theorists are interestingly

developed and are related to many problem areas. This chapter is an important contribution but is perhaps of more value for general psychology than for social psychology. W. W. Lambert had the even more difficult task of relating stimulus-response theory to social psychology. His chapter is better on the presentation of stimulus-response theory than in dealing with the relationship to social psychology. The reasons for the decline of behavioristic theory in social psychology are not faced. One reason for the decline lies in our inability to define clearly the stimulus in social situations. Until we can do this, reinforcement theory will be able to give us explanations only after the fact. Social psychology desperately needs a new objective approach, but Lambert's chapter is an indication that it can not come from the old stimulus-response psychology with its atomism and its grossly oversimplified models.

The most difficult assignment in the theory section was that given to T. R. Sarbin on role theory. While this is a key area for social psychology in relating psychological processes to social structure, there is a dearth of good material on the subject. Sarbin is rightfully concerned with the relation between role and self, but he gives little attention to the relationships existing in a system of roles. It is not very fruitful to consider roles in isolation, for the virtue of the concept is that it takes us beyond the individual into the social system which gives the role its meaning. The concluding statement of the chapter postulates variation in role enactment as a function of at least three variables: validity of role perception, skill in role enactment, and the current organization of the self. This expresses well the orientation of the author and his preoccupation with the pe-

ripheral problems of role theory. Nothing is said about the demand character of the role, the sanctions employed in the role system, the function which the role has for the whole system, the competing nature of other roles, the nature of the collective structure of which the role is a part, etc.

The section on research methods is not a complete coverage of methodological and measurement problems, but most of the major areas are included and the treatment represents a high level of competence. The introduction by Allen Edwards on "Experiments: Their Planning and Execution" gives the proper anchorage in experimental design and procedures. F. Mosteller and R. R. Bush follow with an account of selected quantitative techniques which focuses on the needs of social psychologists and which emphasizes non-parametric methods. Just as we require all students to take courses in statistics, so too should we require writers of statistical texts for these prescribed courses to read the chapter by Mosteller and Bush to see how quantitative techniques can be effectively communicated. The chapter on attitude measurement by B. F. Green is a valuable addition to this section and summarizes the assumptions, procedures, and uses of the methods and models in this field. As in the preceding chapter the author is down to earth in his common-sense judgments about the developments in attitude measurement. The two chapters on data collection "Systematic Observational Techniques" by R. W. Heyns and R. Lippitt and the "Interview: A Tool of Social Science" by E. Maccoby and N. Maccoby are nice combinations of theoretical analyses and practical considerations of the researcher's problems. The chapter by the Maccobys is note-

worthy in cutting through the verbose controversies about techniques and in dealing with the essential dimensions of the interviewing problem. It is valuable too for its reporting of methodological research. The chapter on observational techniques will be very useful because of its inclusion of many different types of observational scales. As the authors suggest, we need to have more investigators utilizing the same observational system if we are to build knowledge in our field. The growing popularity of the sociometric technique and its applicability to social interaction and social structure are recognized in a special chapter on "Sociometric Measurement" by G. Lindzey and E. F. Borgatta. The centrality of the technique to social psychology is emphasized and developments both on the data collection side and on data analysis are described. The authors felt the necessity of choosing between a concept of sociometric technique which followed closely Moreno's early work and a broad concept of the term equivalent to the measurement of social behavior. In choosing the former, they neglected the alternative of accepting the basic technique of reciprocal preferential naming or ranking in the Moreno tradition, but of applying it to a much wider range of problems than personal attractiveness, leadership, and informal patterns in small groups. It is of interest that as this chapter was being written there was already under way a research program carrying the applications of the sociometric technique into new areas, namely, the identification of organizational structure and the processes of communication and control within this structure (the studies of E. Jacobson, S. Seashore, and R. Weiss).

B. Berelson's book on *Content Anal-*

ysis in Communication is neatly condensed by its author to describe the procedures, uses, and problems of qualitative analysis. The weakest chapter in the methods section is the one on cross-cultural method. There is little consideration of the methodological problems of dealing with secondary rather than primary data and little attention to the possibilities of controls in the collection of primary data. In all fairness, however, this chapter is representative of the state of the field. Fortunately, the future is more promising in that cross-national research has led to methodological thinking and there is now in preparation a Unesco-sponsored report by S. Rokkan who, with H. C. J. Duyker, has already outlined some of the dimensions of the problem in the *Journal of Social Issues*, Vol. 10, No. 4, 1954.

The final chapter in the research methods section by D. O. Hebb and W. R. Thompson is on the "Social Significance of Animal Studies." Paradoxically, it has as much, if not more, to offer to the social psychologist as any of the chapters in the two volumes in its wise counsel, in its perspective, and in its insights. It affords the stimulation one would expect from the chapter on "Culture and Behavior." The social psychologist can study with profit the development of different order concepts for the description of animal behavior—descriptive categories which grow out of intensive preoccupation with an understanding of the phenomena being investigated. In brief, this chapter on animal studies is a strong argument for the inclusion of a biologically oriented approach to the problems of social psychology.

The second volume deals with social interaction and applied social psychology. Chapter 18, "Socialization" by I. L. Child, and Chapter 26,

"National Character" by Alex Inkeles and D. J. Levinson, are particularly impressive. Both of these accounts are thoughtful in their organization and interpretation and resourceful in their suggestions for future research. Child, without sacrificing rigorous standards, faces fully the contradictory findings and the weakness of the research methods in some of the studies, and still makes maximum sense of his material. Research on the socialization process is in a better position for advances because of this chapter. Inkeles and Levinson, in their brilliant account, develop fully the implications of national character both as an independent and as a dependent variable in social psychology. This is a strong chapter methodologically in its specifications for common, standard empirical categories of observation, and in its promising framework for future research. Moreover, it comes closer than any other chapter in adequacy of treatment of social structure variables.

One of the classic chapters in the original *Handbook of Social Psychology*, edited by C. Murchison, was Dashiell's summary of the experimental studies of group effects. This topic is now covered by two chapters which ably represent the new conceptualization of group variables and the new types of group experiments—Chapter 21, "Experimental Studies of Group Problem Solving and Process" by H. H. Kelley and J. W. Thibaut, and Chapter 22, "Psychological Aspects of Social Structure" by H. W. Riecken and G. C. Homans. The work of Kelley and Thibaut is a substantial contribution and, although it neglects some historical origins, is the most comprehensive and understanding account of the experimental literature available. The piece by Riecken and Homans

is limited but ingenious. It is an interesting attempt to reduce group activity to a few simple principles, most of which are generalizations from individual psychology. For example, the authors offer the hypotheses: (a) a member of a group chooses or likes another member to the degree to which the latter's activities realize the former's norms and values, (b) the more one member likes another the more he will originate interaction with him, (c) the liking of one member for another is the product of the amount of interaction and the importance of the second member's activities in terms of some group norm. Basic to much of the thinking in this chapter is the old doctrine of self-consistency. Granted the validity of the concept, many social psychologists may want to know how much can be expected from this concept to help in an understanding of group process and group structure.

In keeping with sound tradition there is the usual anthropological piece on culture and behavior. This time it is presented on a grander scale than usual, albeit in somewhat verbose fashion. It is essential that social psychologists should be constantly reminded of how culture-bound their assumptions, concepts, and methods can be. At some point, however, there should be hope of systematic substantive findings in this area as well as the expression of a point of view. C. Kluckhohn has taken a forward step in not only admitting the possibility of universal behaviors but of urging their study. The theoretical position of the article is difficult for the psychologist to follow. On the one hand we are told that anthropologists and psychologists start from data of the same order, namely, human behavior, and that

they may even start from the same particular data. The differences come at the level of abstraction and conceptualization. This is a clear and accurate statement. But then we are told that we should consistently and explicitly recognize the interdependence of cultural and psychological phenomena. Here is the source of the real confusion. If anthropology is on a different level of interpretation than psychology of the same phenomena, then we can not talk about the interdependence of the phenomena. If one observer conceptualizes the flight of soldiers in battle as the ascendance of fear motivation and another sees it as the decline of an empire, we can not put these interpretations into the same system of data as interdependent variables. This confusion between the level of data and the level of concepts limits the chapter in its attempts at integrating culture and behavior.

The section on applied social psychology covers four topics: prejudice and ethnic relations, effects of mass media, industrial social psychology, and the psychology of voting. These chapters maintain the high level of reporting and organization of the *Handbook*. It is most helpful to have an article on industrial psychology which sees the problems as field problems in social psychology, as does M. Haire's sophisticated and stimulating presentation. Haire rightly fingers social organization as the focal point in industrial psychology, but he does not take full advantages of the studies and thinking of the workers in this field (Barnard, Simon, Whyte, Worthy, etc.). The Hovland chapter on mass media is a welcome addition to social psychology. It brings together material which otherwise might escape the social psychologist. It is well organized and research ori-

ented in approach so that the interested reader is stimulated to do experiments to fill the indicated gaps.

There are two major omissions in the comprehensive list of chapters, and even these may be considered a reflection of the present state of our discipline rather than an error of editorial planning. The first major lack is the absence of any systematic treatment of social-structure variables. That small face-to-face groups affect behavior is recognized and at the other extreme there is a chapter on culture and behavior. But between these extremes there is no account of the major forces affecting the lives of people save as they are peripherally treated in chapters devoted to some other topic. The most elementary fact about any society is that behavioral relationships are structured and patterned in the form of social organizations, institutions, and groups. Power is exercised as a function of existing social systems, and people live, struggle, fail, and prosper in relation to the social structure of which they are a part. Social behavior does not occur in a vacuum but even the chapter on role behavior treats social roles as if they existed apart from the arena of group struggle and group functioning. The chapter which might be expected to deal with this material, "Psychological Aspects of Social Structure," is concerned only with small groups, and its attempts at general principles take no account of the nature of social structure. Social class, stratification, the exercise of power, the analysis of institutions, the mechanisms of social control, social movements, and change are neglected. This neglect is in part due to the heavy phenomenological bias of present-day social psychology and, correspondingly, of the authors.

The second major omission is the failure to include a consideration of the relationship between personality variables and social variables. Many of the pages are devoted to reducing social problems to personality dimensions. Many group effects, for example, are reduced to the individual's striving for self-consistency. The fit between personality type and social structure is described in the chapter on national character, but other aspects of the interaction effects of personality and social variables are not explored. In other words, we have perpetuated the old dichotomy of approaches: either all individuals are affected similarly by group conditions or all group effects are explained as the expression of personality mechanisms. If social psychology has any unique subject matter, it may well lie in this neglected area of the interaction effects of personality and social settings.

The publication of a handbook makes appropriate some observations about the development of a field in more global terms than is ordinarily the case. When the handbook represents fairly accurately, as does the present one, the principal areas of effort and the major theoretical approaches, this assessment is greatly facilitated. If one were to compare present-day social psychology with the state of the field when the Murchison handbook appeared in 1935, many changes would be obvious. In the first place there has been a marked shift from a behavioristic approach to field theory and to Freudian theory. Perception and cognition have become central topics for many investigators, and where there is an attempt to go beyond phenomenology, the recourse is to Freudian mechanisms and dynamics. In the second place there has been a growth

of techniques and methods which are appropriate to social-psychological problems. There has been a corresponding increase in the quantity and quality of research and experimentation. Moreover, social-psychological research has on numerous occasions become a collective enterprise with many large-scale projects utilizing the cooperative efforts of numerous workers. In fact, some of the chapters in this handbook have involved a group of colleagues in their preparation. Then too, many new subject matter areas have been investigated, including political and economic behavior.

Somewhat less obvious among the changes is the origin of research and experimentation. The older social psychology started from everyday problems and attempted to account for the variance of behavior in these situations; experiments were more the miniature of the practical problem, as in the work on prestige suggestion or the accuracy of jury judgments. Today these practical interests continue, but in addition the springboard for research is more often an hypothesis based upon some theoretical insight or attempt at derivation from a theoretical model. This is undoubtedly a healthy sign in the growth of a science. But in the effort to avoid framing a problem in applied terms, it is not necessary to divorce theoretical formulation from the data of our field. The chapter by Hebb and Thompson suggests the importance of intensive concentration upon the phenomena under study. The current trend away from rank empiricism is sound, but it can lead to an overemphasis upon exercises in formal logic at the expense of discovery of significant relationships in social phenomena. Another cause for concern in present trends is the lack of

thorough and sustained exploration of a related set of problems so that we can accumulate knowledge in an area and be in a position to make definitive statements about it. Thus far too much effort has been expended at defining or redefining a variable and demonstrating that it is operative at something better than a chance level.

Finally, what is still needed is a theoretical system which could handle the problems of social relationships and structure, not as they happen to be reflected in the psychological field of the individual, but as systems of events in the social field.

Received February 1, 1955

Dr. J. H. G. M. van der Stoep, Department of Psychology, University of Pretoria, South Africa.

Dr. J. H. G. M. van der Stoep, Department of Psychology, University of Pretoria, South Africa.

Dr. J. H. G. M. van der Stoep, Department of Psychology, University of Pretoria, South Africa.

Dr. J. H. G. M. van der Stoep, Department of Psychology, University of Pretoria, South Africa.

Dr. J. H. G. M. van der Stoep, Department of Psychology, University of Pretoria, South Africa.

BOOK REVIEWS

HOCH, PAUL H., & ZUBIN, JOSEPH. (Eds.) *Depression*. New York: Grune and Stratton, 1954. Pp. ix+277. \$5.50.

Perhaps the main value of this book is that it brings together a wide assortment of material related to depressions and allied conditions. It consists of 16 independent papers originally presented at the 1952 meeting of the American Psychopathological Society. Represented in the symposium are theories and findings from the fields of clinical psychiatry, psychodynamics, anthropology, endocrinology, biochemistry, biometrics, and hospital administration. Attention is given to depressive reactions in children, the aged, cancer patients, and soldiers exposed to isolated Arctic conditions. Most of the articles are reviews of previously published work. As might be expected from a symposium of this nature the individual chapters vary markedly in quality. Kallmann's analysis of genetics in manic-depressive psychosis, Zubin's review of biometric methods in psychopathology, and Steinbrook's cross-cultural evaluation of depressive reactions are outstanding. Some useful criteria for the differential diagnosis of manic-depressive patients are contained in the chapter by Lewis and Piotrowski. Kalinowsky's enthusiastic defense of electric convulsive therapy conveniently ignores the well-established finding—mentioned in an earlier paper—that the present irrational methods of shock therapy, at best, simply hasten the improvement of those who would improve anyway. It may be true, as maintained by Spitz, that institutionalized infants deprived for several months of emotional supplies normally provided by

the mother undergo progressive developmental decline, but the absence of adequate experimental controls leaves room for doubt. Might it not be that infants kept in institutions for long periods are biologically less favored than infants who are adopted or placed in foster homes after a brief stay in an institution?

Much still remains to be learned about depressions but in focusing attention on and in providing a wide-range view of the problem the APS has performed a useful service.

JAMES D. PAGE

Temple University

SAVAGE, LEONARD J. *The foundations of statistics*. New York: Wiley, 1954. Pp. xv+294. \$6.00.

Only the title of this provocative book will be easy reading for most psychologists. By "foundations of statistics" the author mainly refers to the axiomatic and logical bases of probability theory. After a brief introductory chapter which notes three major points of view about the interpretation of probability (objectivistic, personalistic, and necessary), a personalistic or subjectivistic theory of decision in the face of uncertainty is developed in formal manner involving an assortment of definitions, postulates, theorems, lemmas, and corollaries. The theory is shown to imply the construct of utility in economic behavior.

The second half of the book is devoted to the impact of personalistic probability on minimax theory, the theory of games, point estimation, interval estimation, and the testing of statistical hypotheses. Overconscientious appliers of common procedures in statistical inference may be somewhat traumatized by the

large component of arbitrariness, vagueness, and crudity attributed to the rationale of these procedures.

LEONARD S. KOGAN

*Institute of Welfare Research,
Community Service Society of
New York*

ROBACK, A. A. *Destiny and motivation in language*. Cambridge, Mass.: Sci-Art, 1954. Pp. 474. \$10.00.

This is a book of many ideas about language. It will be read with more profit by the layman than by the scholar or scientist because it is written by an amateur (using this word in its exact sense) rather than by an expert philologist or linguist.

Dr. Roback has been interested in language since the age of five when he asked his teacher whether the P'rizzim were the same as Parisians. Had it not been for rebuffs from his language professors at McGill and from Münsterberg at Harvard, Roback might have become a philologist or a researcher in the field of psycholinguistics. In writing this book the author has returned to an early intellectual passion. Appropriately, the book bears the imprint of adolescent exhibitionism and enthusiasm for ideas more than it does the mark of mature and discriminating scholarship. I found it delightful reading.

Destiny and Motivation in Language is a collection of 22 essays on a wide variety of subjects relating to language. I like particularly the chapters on destiny in names, shmoo and shmo, and the psychology of slang. Brutus was destined to be a brute, Boccaccio to be foul mouthed, Kierkegaard to be a gloomy graveyard existentialist, Hitler (Schicklgruber) to be a grave sender, Sir Russell Brain to be an authority on the brain, and the writer of this review to be bald since Calvin means

bald. (He is bald!) This is the sort of whimsical, tongue-in-cheek psychologizing that people engage in after a few drinks and rarely take seriously. Perhaps they should!

"Shmoo and Shmo" is a charming essay on the linguistic connection between the name given by Al Capp to his fabulous little cornucopian creatures and the feminine form (shmooe) of the Yiddish word for the erect penis (shmuck). Shmuck is a hard sounding word; shmooe, shmoo, and shmo are soft sounding. This in essence is Roback's vocosensory theory of the origin of language. The muscular pattern made by the mouth plus the sensory attributes of saying the word contain the meaning of the word. Words sound and feel like the things they refer to. Shmuck sounds and feels like an erect penis, shmoo like a soft comforting vagina. Al Capp (born Caplin) unconsciously selected the feminine form of a common Yiddish word, which he must have heard as a child, as the name for his supernurturant birds. This essay reminds me of the perspicacious writings of Kenneth Burke on the relation of letter and word sounds to meaning. (How many psychologists are familiar with Kenneth Burke's game with words which he calls "joycing" after the greatest modern wordmaster of them all?)

In my opinion, the most valuable chapter in Roback's book is the one on slang. In this essay, much of which was written forty years ago, Roback manages to sustain a line of thought long enough to reach some pretty interesting conclusions. One such conclusion is that slang is the rebel's form of expression, whether the rebel be an adolescent, a criminal, or a member of a persecuted minority.

Although this book is not an important contribution to the psychology of language, it will be read with

interest by people who like the stimulus of fresh, off-beat speculation. Dr. Roback has a keen eye for the ironical, an affectionate feeling for words, and a highly permissive attitude toward ideas. It is pretty apparent that *Destiny and Motivation in Language* was undertaken as a labor of love. It deserves to be read and criticized with the same emotion.

CALVIN S. HALL

Western Reserve University

COOPER, LINN F., & ERICKSON, MILTON H. *Time distortion in hypnosis*. Baltimore: Williams & Wilkins, 1954. Pp. ix+191. \$4.00.

In this provocative little book the authors present the results of various investigations of hypnotically induced distortions of time perception and their clinical applications. Those of the readers who have not had the opportunity to read Cooper's earlier publications on this topic will find this book to be, for the most part, pleasant and fascinating reading. For those who are familiar with the aforementioned papers this monograph will be somewhat disappointing since the greater portion of its bulk duplicates the earlier reports. Nevertheless, a fair amount of previously unpublished experimental data has been added and the clinical material is entirely new. In addition, the material dealing with basic concepts and methodology represents an important expansion of the earlier work, and this should prove valuable to experimenters interested in this area.

Essentially, the subject matter deals with slowing down a person's perceived or subjective time through the use of hypnotic suggestions and the effects upon various mental functions. Data are presented which indicate that some functions may be speeded up. It is to be regretted that Cooper and Erickson have preferred

to cover a relatively wide range of functions in an exploratory manner rather than limit themselves to a lesser number of more definitive studies. The consequence has been that, although the reported data are superficially convincing, closer inspection tends to leave us wishing for something more substantial. There is a serious lack of objective measures and quantified data, and many of the reported experiments tend to be methodologically oversimplified. The book would also have been more satisfactory if the authors had made a greater attempt to relate their work to current psychology. On the other hand, it should be kept in mind that the topic is admittedly a difficult one to deal with experimentally, and that Cooper and Erickson are essentially pioneering in a new area of investigation. Taken as a totality the reported data do not permit us to categorically ignore hypnotic time distortion and further investigation is clearly warranted.

Cooper devotes some space to theorizing about a general mechanism for the action of hypnotic suggestions. In a somewhat laborious and circuitous manner he reaches the conclusion that hypnosis imbues words with a particular effectiveness to evoke specified experiences via their associated "meaning-tone." This is also the mechanism responsible for time distortion. In this reviewer's opinion, this conclusion hardly seems to elucidate the main question and leaves many other questions unanswered.

In a sense the clinical material reported by Erickson may be the most convincing evidence for the reality of time distortion. The use of the latter seems to be potentially promising for brief psychotherapy and for the treatment of obdurate cases. It also brings to the fore certain temporal aspects of the process of therapy

which have been largely ignored to date.

In final evaluation, the reviewer believes that, in spite of certain shortcomings and its rather specialized nature, this monograph will be of interest to a wide group of readers. It represents an original attack upon the problem of temporal perception which cuts across major areas of psychology. The experimenter will discover in it both a new tool and a rich field for further research. Theoreticians, particularly those with a philosophical turn of mind will find food for thought and for speculations. Finally, it gives the clinician and therapist an intriguing new approach to the problems of psychotherapy.

ANDRÉ M. WEITZENHOFFER

University of Michigan

McKEON, RICHARD, et al. *Interrelations of cultures*. New York: Columbia Univer. Press (Unesco), 1953. Pp. 387. \$2.50.

This little book deals with eight cultures, particularly their historical origins and value systems. "Interrelations" are treated mainly in historical terms except in McKeon's introductory essay and the final brief statement by a Unesco committee. The authors are predominantly historians and philosophers, and these points of view are central in the discussion. Indian "psychology" is given a few pages (pp. 134 ff.), but they will not appeal to most academic psychologists in the United States. Griaule makes some very suggestive remarks about cognition and learning among African Negroes. There are some excellent—and to this reviewer quite fresh and new—treatments of Indian, Chinese, and African Negro esthetics. Anthropology enters principally into the consideration of African cultures; it gets short shrift on Japan (p. 121).

As always with volumes of this sort, the contributions vary greatly in interest and quality. The two pieces on American culture are extremely thin. On the other hand, Zea on Spanish-American culture is profound and provocative. Atreya on India gives a more compact and intelligible account of central themes than the reviewer has ever previously encountered. Chatterji's characterization of Indian culture also has force and point, though one feels he strains to present it in a manner that will fit best with certain currents in Western thought.

In general, this volume will attract only those psychologists who are interested in the cross-cultural study of values or those who enjoy somewhat discursive essays seasoned with occasionally exotic facts. Yet the best of these chapters are extraordinarily thoughtful syntheses of vast bodies of material, and Unesco deserves thanks for making them available.

CLYDE KLUCKHOHN

Center for Advanced Study in the Behavioral Sciences

ZISKIND, EUGENE. *Psychophysiological medicine*. Philadelphia: Lea and Febiger, 1954. Pp. 370. \$7.00.

Here is a book written for the general medical practitioner that will interest psychologists more by its purpose than by its content. Its premise is that one-third of the patients seeking medical aid have symptoms solely or primarily due to psychogenic pathology and another third combine psychogenic with physiogenic pathology. The requirement for psychodiagnosis and psychotherapy in the treatment of general medical disorders is clear indeed.

Dr. Ziskind, head of the department of psychiatry and neurology at the Cedars of Lebanon Hospital in Los Angeles, has been coordinating a

training program for physicians to meet this need. He drew heavily on this experience in preparing the portions of this book dealing with diagnosis and therapy.

The book consists of four parts. The first one provides background and orientation to the problem. The second part is concerned with diagnostic and psychotherapeutic technique. The third part consists of a review of familial and other social factors in psychopathology, plus a general treatment of psychiatric syndromes and common psychiatric emergencies with which the physician may be confronted. The fourth part includes a useful review of contemporary schools of psychiatric thought.

The writing style is easy and lucid. In keeping with the aim of the book, there is an almost "do it yourself" quality to it. This is one of its shortcomings. Skill, personality, and character requirements for effective psychotherapy are stated as precepts, as if the reading by the physician is a sufficient fulfillment of the requirements. The requirements themselves are not at issue. Ziskind states that the personality and attitude of the therapist must be encouraging. He should be kindly, sympathetic and interested, permissively nonmoralistic, nonjudgmental, and nonauthoritarian (p. 149).

Within the framework of a supervised training program at a hospital there is some modest assurance that these requirements may be met. As a minimum, a physician lacking these characteristics can be discouraged from further participation. Ziskind reports that the hospital program has been highly successful, and it may be assumed that the success is due in no small measure to the availability of a well-trained psychiatric staff.

For a training program of this type

this book would be a useful systematic guide. However, the book is addressed to the practitioner at large. The wisdom of its use independently of supervision will be seriously questioned. Ziskind believes that because psychophysiologic illness is so rampant, the physician is obliged to deal with it as best he can anyway and after reading this book the physician can approach this task with less anxiety.

What is at issue is whether the patient is better off if he is exposed to the risk of therapeutic mismanagement. This is the dilemma posed by Ziskind's effort. The merits of his effort are so obvious that it would be highly valuable to assess the magnitude of the risk empirically even though this may be a large undertaking.

Against such data as a background, other solutions may also be considered. These would include the overdue modification of medical school curricula and the establishment of a more cooperative working relationship between the medical and psychological professions.

JOSEPH E. BARMACK

College of the City of New York

MURPHY, G. & BACHRACH, ARTHUR J. *An outline of abnormal psychology*. (Rev. Ed.) New York: Random House, 1954. Pp. ix+597. \$1.45.

This is a revision of Dr. Murphy's 1928 contribution to The Modern Library Series and almost everything but the title has been changed. Dr. Murphy gives most of the credit for the selection of the material to his co-editor. The first chapter is a Gardner Murphy original. In addition to chapters on the dynamics of schizophrenia, psychoneuroses, and paranoid disorders, there are chapters on specific topics such as adolescent

conflicts and adjustment, obesity, the addictive drinker, psychological factors in epilepsy, the individual and society, Rorschach variables and ESP scores, mongolism, and the treatment of the withdrawn child. The editors have selected published material which has breadth, reader importance, technical accuracy, and simplicity of report. The unity of the book is striking, in view of the wide range of subject matter and the sources of the papers included. An appropriate and helpful glossary is appended. The volume would have been improved by an index, especially since it is suitable for college classes.

This is not just another book on abnormal psychology, as the title may suggest, but a collection of readings which may be used either to supplement existing texts, or alone, as an aid to stimulating insight and understanding in the searching layman. As psychology's role in the community continues to be enlarged, there will be more and more need for authentic, readable, and relevant books such as this for the average reader.

FRED MCKINNEY

University of Missouri

HOIJER, HARRY. *Language in culture*. Chicago: Univer. of Chicago Press, 1954. Pp. xi+286. \$4.50.

The late Benjamin Whorf's dramatic presentation has interested a large part of the intellectual world in the hypothesis of "linguistic relativity." Whorf maintained that the cognitive style of a people is in correlation with, and perhaps determined by, the lexicon and structure of the language spoken by that people. In 1953, at the University of Chicago, a group of scholars representing linguistics, anthropology, psychology, and philosophy discussed for one week the

"relativity" thesis of Whorf. The report of the conference comprises seven papers and nine discussion sessions.

With Whorf, as with Freud, no reading, however careful, can tell us exactly what the author meant if exact meaning implies operational definition. The exegesis of Whorf is a principal topic of the conference. This portion of the discussion may seem to be wasted since the problem of Whorf's exact intentions is, after all, a rather trivial historical question. In fact, however, each discussant takes this opportunity to propound his own views and receive criticism of them, letting Whorf absorb the odium for whatever may be said that cannot survive criticism.

Whichever of several things Whorf meant to say the conference members generally agree that his data (largely from his study of Hopi) are suggestive rather than conclusive. "Linguistic relativity" is a collection of hypotheses about language, culture, and cognition. The conference accordingly gives some time to the design of research intended to test these hypotheses. Hoijer makes an interesting proposal for comparing general cultural similarities with linguistic affiliations among the Navaho, Hopi, and other American Indians. Hockett gives the best feeling for method and also provides valuable examples in his comparison of certain limited features of English and Chinese. In a final discussion research ideas are rather carelessly tossed out. Several of these are promising (notably those of Lounsbury and Kaplan) but the reader is disappointed by the failure of the conference to think through any one idea to a point where its practicability and merit can be judged.

Hoijer has condensed and edited

some 919 pages of typescript to obtain the 150 pages of discussion that appear in the book. These discussions read smoothly and are generally intelligible. However, the transcribed dialogue of ideas never has the sustained direction and thoughtfulness of a literary construction. The unfortunate fact is that not even every other idea expressed in a conference is worth recording. As a result this book makes an extremely uneconomical presentation of a small number of good ideas.

ROGER W. BROWN

Harvard University

KLOPPER, B., AINSWORTH, M. D., KLOPPER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. Vol. I. *Technique and theory*. Yonkers, New York: World Book, 1954. Pp. x+726. \$6.50.

There is something for everyone in this first volume of a two-volume work on the Rorschach. Probably the primary use of the book will be as a teaching manual for students learning the mechanics of scoring and the feeling for interpretation, but even experienced users of the technique may read with profit the sections on theory and interpretation.

The sections on administration and scoring are written clearly and numerous examples are provided to clarify scoring problems. Few changes in the scoring system presented twelve years ago in *The Rorschach Technique* have found their way into the present procedure. There is some clarification of methods of scoring shading responses though these distinctions are likely to continue to be frustrating to students. Ten popular responses are listed—the same ones as twelve years ago, and a slightly different list from that offered by other experts. The form-level rating procedure, an attempt to give a numerical score for each percept reflect-

ing its accuracy, will continue to baffle statisticians and others with a bias against adding, subtracting, and averaging cabbages and kings.

The sections on interpretation are well written and will stir little controversy among experienced users of the technique. Some readers are certain, however, to criticize the failure of the authors to support their statements with data in the literature. The authors, anticipating this objection, present their interpretations as "hypotheses" and indicate that the usefulness of these hypotheses has been established in clinical practice and that they may be used "pending the conclusion of the extensive investigations that will be necessary to evaluate their relative validity." Apparently existing validation studies do not fit their criteria. Certainly clinical use has resulted in few changes in interpretation. The authors also indicate, wisely, that "the integration of the findings into some sort of meaningful dynamic picture depends largely upon the examiner's basic understanding of human personality." Again, there is much in the literature which could be used to document this hypothesis.

The section on validation problems makes a case for the point of view that "there is no sharp dividing line between validation research and the clinical use of the Rorschach technique . . . indeed, the technique itself is continually changing through the validation process." A test of this latter hypothesis through a comparison of scoring and interpretation procedures by the same authors over the years, and by an examination of possible gradual rapprochement between different schools, would possibly raise some doubts here. The section on validation concludes with the statement that "validation research should be planned and the results discussed by those who have more than

a superficial understanding of the basic concepts involved in Rorschach interpretative hypothesis. . . . " Few will disagree, though the statement leaves a nice opening for a rejoinder by the statisticians!

To single out these statements does not give an accurate picture of the generally excellent discussion of the problems of validation and of existing approaches to these problems.

Also included is a brief, though stimulating chapter on implications of contemporary personality theories, a somewhat more esoteric chapter on ego psychology, a long case study, and a Rorschach Prognostic Rating Scale (experimental) designed to predict a patient's response to psychotherapy. This latter *may* be just what Eysenck has been looking for, though again it may not. It would be interesting to hear him discuss it with the authors at a symposium. At the end of the book, as an appendix, is an added chapter on Freudian and Jungian approaches to the development of ego and self. The author is Jack Fox, who is not represented on the masthead, perhaps because there is no mention of the Rorschach in the chapter.

The book contains no general list of references though there is a bibliography at the end of many chapters. Those interested in or favoring particular cards of the Rorschach will find a list of the pages on which each card is mentioned.

In a work of this scope it is easy to find statements or points of view with which one can disagree. Also with a second volume in the offing one may find his objections concerning lack of some form of content answered with the companion volume. There is little doubt that the present work will find a wide audience and deservingly so.

GEORGE W. ALBEE

Western Reserve University

KATZ, D., CARTWRIGHT, D., ELDERSVELD, S., & LEE A. MCC. (Eds.) *Public opinion and propaganda*. New York: Dryden, 1954. Pp. xx+779. \$6.25.

The Society for the Psychological Study of Social Issues has produced an excellent book of readings on public opinion and propaganda. In keeping with its interdisciplinary orientation, a wide range of the relevant literature in political science, history, anthropology, sociology, economics, and psychology has been culled to provide a selection of readings (74 in all) showing the possibilities of the empirical approach to problems in the area as well as the societal context, the political structure, and the social-psychological dynamics of opinion formation. There is as wide a coverage of points of view, types of studies, and problems, both practical and theoretical, as one has a right to anticipate in a book of readings intended for courses in public opinion and propaganda. On the basis of the sampled advanced work included in the collection, one wonders, however, whether the general area of public opinion and propaganda will acquire that eventual "scientific" status which the editors suggest such work may adumbrate. "Control" and "number" are not enough. After all, Ptolemaic theory also had its systems of quantification.

IVAN D. LONDON

Brooklyn College

MCMILLAN, B., GRANT, D. A., FITTS, P. M., FRICK, F. C., MCCULLOCH, W. S., MILLER, G. A., & BROSN, H. W. *Current trends in information theory*. Pittsburgh: Univer. of Pittsburgh Press, 1953. Pp. xiii+188. \$4.00.

This is the seventh in a series of annual conferences on *Current Trends in Psychology* arranged by the Depart-

ment of Psychology of the University of Pittsburgh. The seven participants and the titles of their contributions are: McMillan, "Mathematical Aspects of Information Theory"; Grant, "Information Theory and the Discrimination of Sequences in Stimulus Events"; Fitts, "The Influence of Response Coding on Performance in Motor Tasks"; Frick, "Some Perceptual Problems from the Point of View of Information Theory"; McCulloch, "Information in the Head"; Miller, "Information Theory and the Study of Speech"; and Brosin, "Information Theory and Clinical Medicine (Psychiatry)."

As is frequently the case with such compendia, this one is rather a hodgepodge with different authors trying to do different things in different ways. The unifying theme, such as it is, is modern communication theory. McMillan attempts a brief and general account of communication theory, in which attempt he is very successful. It is gratifying to know that mathematicians *can* communicate with psychologists if they want to.

Grant talks about a number of experiments dealing generally with partial reinforcement and the discrimination of stimulus frequencies occurring with various probabilities. Although the experiments are interesting in their own right, the attempt to relate the results to information theory is strained. It is hard to escape the feeling that it would have been much less cumbersome to talk about these studies in other terms.

Fitts is concerned with the speed and accuracy of discrete and continuous movements and with stimulus-response compatibility. He reports some exceedingly ingenious experiments but it is difficult to know how much their conception is in debt to information theory. At any rate, it is possible to argue that the principal

contribution of information theory to these results has been to suggest the use of logarithms.

McCulloch has an interesting survey of some neurophysiological problems of sensation and perception and Brosin talks about some problems in clinical psychology and psychiatry. In both cases the principal connection with information theory has been the translation of some common psychological and physiological terms into some uncommon ones, e.g., "quantize," "modulate," and the like. If both authors had applied terms ordinarily used in their respective fields, their discussions would undoubtedly have contained as much information and would probably have been better communications.

Of all the authors only Frick and Miller describe some classes of psychological experiments to which information theory has been able to contribute something other than a new vocabulary. Miller is especially levelheaded in his appraisal of the usefulness of information theory for psychology and a paraphrase of the first two pages of his contribution might well summarize the outcome of this conference: There are not yet many places in the study of human behavior where information theory can be profitably applied.

Since the chances are good that you do not agree with the reviewer's appraisal of information theory, you ought to read the book yourself. It is worth reading.

A. CHAPANIS

The Johns Hopkins University

ZILBOORG, G. *The psychology of the criminal act and punishment*. New York: Harcourt, Brace, 1954. Pp. xi+141. \$3.50.

This short book grew out of a Yale lectureship under the joint auspices of schools of law and medicine. It

examines the motivation and consequences of punishment of the criminal and points to the "curative powers of healthy self-punishment which never becomes hostility."

It will make its major contribution to the practicing professions, but psychologists acquainted with the hypotheses reviewed here will welcome their emphasis by a writer who is well regarded in his own and related fields. Zilboorg presents excellent suggested courtroom ethics for the representatives of the behavior sciences.

FRED MCKINNEY

University of Missouri

WOLSTEIN, BENJAMIN. *Transference: its meaning and function in psychoanalytic therapy*. New York: Grune & Stratton, 1954. Pp. xiii+206. \$5.00.

Trained in philosophy, psychology, and Sullivanian analysis, Wolstein has given thought to the rich source of fantastic data brought to light chiefly by methods stemming from Freud: dreams, free association, "transference" observations—areas largely avoided by experimentalists. He also understands operationism and precise models of science. His book demonstrates well the feasibility of bringing into scientific consciousness those analytic concepts which sometimes seem to have come up dripping with the amorphous illogic of their dreamworld origin.

His review of Freud's and later analytic thinking suggests that Freud saw transference as infantile behavior toward parents "transferred" to a present situation. Wolstein sees transference in terms of observable here-and-now behavior. Freud pursued relentlessly the reconstructed memory of an event from early infancy which "determined" the neurosis; using first hypnotism,

then strong suggestion, and finally an authoritarian exploitation of "positive transference." Negative transference he battered down as resistance, and countertransference was not admitted to the patient. Wolstein accepts extrapolative reconstruction of infantile events, but also emphasizes later developmental stages. He keeps the focus on present interactions with careful attention to the therapist as an interbehaving organism (countertransference) rather than as a hypnotic authority.

The book reflects the groping trends of opinion developing from therapeutic experience and from increased awareness of the problems of definition. Many questions related to transference are brought closer to the status of answerable questions.

ROY M. HAMLIN

*Western Psychiatric Institute
University of Pittsburgh*

WEVER, E. G., & LAWRENCE, M. *Physiological acoustics*, Princeton: Princeton Univer. Press, 1954. Pp. xii+454. \$10.00.

This new volume from Professor Wever's laboratory can be immediately recommended to two classes of readers: the few who eat and sleep audition, and the many who need a clearly written text to keep them posted on the latest facts plus authoritative comment from a logical mind on the meaning of those facts.

For workers already immersed in the area, the book provides two values. It brings together much of the material Wever and his associates have scattered over scores of papers in the acoustical, otological, and psychological literature; and it also presents new data and fresh comment on old data, both of the very greatest interest.

The plan of the book is to examine the transmission of acoustic energy

from air-borne vibrations clear through the system up to but not including the primary neuron. After the three short introductory chapters, well over two-thirds of the remaining 375 pages is devoted to the system up to but not including the cochlea. This emphasis on the drum, the ossicles, the tympanic muscles, problems of bone conduction, of the relative efficiency of routes to the cochlea, and of clinical problems in sound conduction form a salutary change from a common notion that "auditory theory" concerns only basilar mechanics and associated neurology. There is much besides a dramatic concentration on cochlear patterns to engage the attention of an auditory theorist.

This book makes clear how Professor Wever has focused the full attention of his laboratory for a quarter century on the whole problem of sound conduction. Beginning at the eardrum and proceeding to intracochlear electrodes, every phase of the subject has been touched upon by him in some, often crucial, experiment. The result is a treatise in the best tradition. Certain problems unsolved since Helmholtz have yielded to Wever's experimental technique. A good example is his chapter on the eardrum as a conical lever. He first summarizes Helmholtz's analogue of the catenary as the principle by which the force applied to the malleus is amplified, marshals the anatomical evidence on the elasticity of the radial and circular fibers, and then repeats Helmholtz's crucial experiment with a technique enabling him to move the manubrium of the malleus accurate to .0005 mm. As a result, auditory scientists will never again seriously consider the eardrum as an amplifier.

As an example of the care and logic exercised in Wever and Lawrence's work, one may cite the series of studies on the function of the round window. Although Eduard Weber long ago had the generally correct notion, numerous attempts by several modern laboratories to document this notion had proved ambiguous. It remained for Wever and Lawrence to concentrate their attack on the round window with quite nice techniques of blocking, probing, loading with an air column, with and without ossicles present, and to show in precise quantitative fashion the proportionate contributions of mass and stiffness to the system, as they affect high-tone and low-tone responses. Thus not only is the main function of the round window logically proved, but the details are expressed in quantitative fashion and a rational explanation offered. Finally, the failure of earlier studies to produce conclusive results is very civilly given by inference, and a true explanation for their data is at the reader's hand for the first time.

The medical significance of Wever's decades of work on sound conduction is best appreciated by otologists, to whom he often personally addressed himself and who have taken the opportunity to underwrite some of his efforts.

To psychologists, the features of Wever's work of most significance are his discovery of the cochlear microphonic (with which of course the name of C. W. Bray is always connected as well), and his insistence on the unsatisfactory nature of "place" as the only principle of cochlear analysis. In his book *Theory of Hearing* he fully discussed the reasoning which led him never to neglect the fact that a nerve fiber

always conveys information by its periodicity as well as by its anatomical location, and which led him to assimilate the characteristics of both space and time in a unified theory. The wording here is reminiscent of Harvey Fletcher's space-time theory, but there are real differences. In Wever's view a fast-acting pressure wave sets up a geometric vibration pattern on the basilar membrane, and this pattern is sufficient to serve as the intermediary between the oscillatory pattern of the stapes and the space-time pattern of nerve response. In Fletcher's view, and in that of many investigators today, the pressure wave serves mainly to initiate a second, much slower, traveling wave moving up the basilar membrane much like a vibration along a shaken rope, and the characteristics of this traveling wave determine the neural space-time pattern.

It must be said that direct observation and certain intracochlear multiple-electrode studies seem to support the existence of a relatively slow traveling wave. Nevertheless, Wever and Lawrence offer the possibility of another interpretation, and do not regard the wave as proved beyond question. In addition, they offer a most ingenious and carefully controlled experiment showing that the cochlear pattern is very similar no matter whether sound is introduced by a tube over the oval window or by a tube over a little hole drilled in the apex. It is admittedly most difficult to explain this result by a slow-traveling wave, as indeed it is difficult to explain many of the extremely rapid phenomena of audition with a slow wave of any sort. In these matters, Wever's theory comes off very well.

Many will wonder just how and in

what region Wever would have the frequency principle for low tones give way to the place principle for high tones; Wever will simply state that such questions are properly the subject of experiment not argument. Moreover, it may be that the experiment of introducing sound at the apex is not crucial; one can believe, for example, that the construction of the apical end of the basilar membrane and associated spiral ligament is not conducive to being snapped like a rope, and that the pressure wave beginning at the apex and hitting the oval and round windows within about 20 microseconds, is reflected back up the canals and initiates a slower traveling wave at the usual end of the basilar membrane. Thus a sine wave impressed upon the fluid at opposite ends could well initiate two traveling waves following each other with only a slight phase lag, well within the limits of measurement, and it would be possible to create cancellation effects quite generally throughout the canals, as indeed Wever and Lawrence found.

Suffice it here only to say that the data on the question as to the emergence of the pattern of excitation in the cochlea, the very voluminous data from several laboratories since the publication of his book of five years ago, have not caused Wever to change his mind as to the necessity for a synthesis of both principles, space and time. And in the discussions which are now progressing in the several journals, Wever, by his persistence and elegance in the laboratory and in the literature, makes quite as good a case as anyone.

J. DONALD HARRIS

*U. S. Naval Medical Research
Laboratory,
New London, Connecticut*

HALL, CALVIN S. *A primer of Freudian psychology*. Cleveland: World, 1954. Pp. xii+137. \$2.50.

Hall's small and unpretentious volume is in most respects just what the author and publisher claim it to be—a relatively nontechnical, yet systematic primer of Freud's theory of personality. A second claim—that its material is selected from and reflects Freud's dynamics of normal rather than abnormal psychology—seems to be a bit overdrawn. Although Freud certainly distinguished between normal and abnormal states, it is equally certain that he did not invoke different dynamic constructs to account for each of them. However, the structure of Hall's book is not vitiated by the artificial dichotomy of dynamics which he introduces, nor is its value lessened as a singularly clear, comprehensive, and unadorned account of essential Freudian theory.

In his introductory chapter Hall gives a brief and informal picture of the steps which led Freud from biological research to psychology. Apart from this account in which his admiration for Freud, the Scientist, is frankly explicit, the author's own views are never obtrusive. They are in fact, hardly discernible. This anonymity of "reporting" is one of the many positive features of the Primer. Hall's careful organization of topics, his succinct introduction to each major new one, and his brief but precise summaries should make this book a prototype.

The lay reader will get a reasonably good overview of Freudian theory from the book and will be assisted further by the list of recommended readings furnished him. For the student, the book should provide an excellent frame of reference for further study. The topical list of ref-

erences to Freud's writings will be of real assistance to him.

DAN L. ADLER

San Francisco State College

FAIRBAIRN, W. RONALD D. *An object-relations theory of the personality*. New York: Basic Books, 1954. Pp. xi+312. \$5.00.

In a series of lectures delivered in the 1940's, Fairbairn developed what for him was a renovation of Freud's libido theory and Abraham's "developmental phase" theory. It is these lectures which appear in the first portion of this book; in succeeding sections there appear three case histories which are used to illustrate the author's resynthesis, and a group of miscellaneous papers.

If one were to use the criterion of internal consistency in evaluating the theoretical contribution of the author, he would come off very well indeed. From lecture to lecture, Fairbairn unfolds a program of interpretation which is essentially free from *non sequitur*, and carefully avoids the loose ends which often characterize this method of presentation. The premise is stated in the observation that schizoid states are found in practically all psychopathological cases, and by implication must develop early in life. They occur, therefore, as resultants of varied relationships between the person and objects which he seeks in response to libidinal prompting. Using these "object-relations" as a key concept, Fairbairn proceeds to reorganize classical interpretations of such things as melancholia, impulse-psychology, and the holy trinity of ego, id, and superego.

If one were to evaluate this New Look by other than consistency criteria, however, it is doubtful that the consensus would be so favorable. For many classical psychoanalysts, Fairbairn will probably represent a par-

ticularly repugnant type of heresy in his disavowal of "accepted" basic principles. To the neo-Freudian, his dynamics may appear too far removed from the social world to be considered practicable. Finally, for those who demand of a conceptualization that the referents be clear, and that labels and names not be treated as entities, there will be little incentive to read beyond the first chapter.

Among the miscellaneous papers in the last section there is one which might be read as an object lesson in righteous indignation. Fairbairn argues here that there is strong discrimination against the acceptance of psychoanalysis (and surrealism) in academic curricula, from which it follows that "... the higher the cultural value of a subject in the university curriculum, the less freedom of thought and enquiry is permitted" (p. 253). The scale by which he measures cultural values is not described.

DAN L. ADLER

San Francisco State College

SHELDON, WILLIAM H., in collaboration with DUPERTUIS, C. WESLEY, & McDERMOTT, EUGENE. *Atlas of men*. New York: Harper, 1954. Pp. xvi+357. \$10.00.

This Atlas is organized as a working guide to be used in conjunction with Sheldon's earlier volumes, *The Varieties of Human Physique* and *The Varieties of Temperament*. It contains somatotype photographs of 1,175 men, selected as representative of the 88 somatotypes identified in the total sample of some 45,000 men ranging in age from 18 to 65 years. As the book is primarily a supplement to the earlier volumes, it is necessary to turn to them for detailed descriptions of the criteria for ratings, unless one can learn, by repeated leafing through the photographs, to identify the differentiating

factors between, say, a 2 and a 3 in ectomorphy. Each somatotype, in addition to photographs of several examples at different ages, is presented with (a) a thumbnail sketch, largely in fanciful analogy to an animal "totem" that is intended to characterize both the physique and its personality, (b) a chart showing the relative position of this and similar somatotypes, (c) a description, including the incidence of this type in the population, and the kinds of pathology, both physical and mental, that are most characteristic of it, and (d) a table giving weight for height and age, and a chart of the H/\sqrt{W} index by age, for purposes of evaluating nutritional status according to somatotype.

Appendices include a glossary; a set of tables for distributions of somatotypes, by age, according to the H/\sqrt{W} (which may be used as a guide for identifying somatotypes); a description of equipment and procedures (including a chart for converting weight and height into the H/\sqrt{W} index); and a table giving the case numbers of subjects in the Atlas having varying degrees of "gynandromorphy" and "texture," the two relatively independent variables.

New data on this system of somatotyping are discussed briefly. In a consideration of the sex factor, Sheldon defends his use of the same standards for both sexes, and presents charts showing the differences in distribution of the somatotypes of men and women. The females predominate in the areas of endomorphy-ectomorphy, with no ectomorphic mesomorphs, while the males have a greater variety and spread of types, the mode being high in the mesomorphic component. An *Atlas of Women*, a companion volume to this one, is in preparation. In dealing with the age factor Shel-

don restricts his discussion to the span 18 through 65 years, but we find that an occasional boy under this age, as young as 11 years, has been slipped in among the illustrations. This seems to me a dubious procedure, because body-proportions change so markedly in boys between 11 and 18 or 21 years. Lacking longitudinal data on the men, the author has made use of his subjects' reported weight histories, together with such casual photographs as they could supply, to project the curves backward in time. Each somatotype has a characteristic curve, as some naturally put on more weight than others. These curves seem plausible, but they need verification by actual repeat measures and assessments of the same men as they grow older. There is a discussion of the effect of nutrition, in which Sheldon defends his system against critics who have shown that such things as semistarvation alter the apparent somatotype.

For those who are concerned with the problem of physique and its relations to personality and to disease, this book is provocative and a challenge to further research on the hypotheses and assumptions presented. For those who use the Atlas as a tool or guide for evaluating subjects, diligent study of the many illustrations should contribute to proficiency in somatotyping. However, it would be easy for the uncritical to overvalue the validity of Sheldon's assumptions. Although at the outset he is careful to call his work a "pilot study" and to point out the need for more validating procedures, the descriptions of each somatotype carry the ring of certainty, e.g., of somatotype 1 3 6 he says, "Flycatchers. Frail lightly built birds who yet are fierce-hearted hunters. . . . When found in the men-

tal hospital . . . the 1 3 6 is most often labeled *paranoid schizophrenia*. . . . But if the 1 3 6 is sometimes paranoid he is often heroic. For wherever paranoia dwells, there also is the home of heroic devotion" (page 45).

NANCY BAYLEY

National Institute of Mental Health

HENRY, ANDREW F. & SHORT, JAMES F. JR. *Suicide and homicide*. Glencoe, Ill.: The Free Press, 1954. Pp. 214. \$4.00.

Subtitled "Some Economic, Sociological and Psychological Aspects of Aggression," this book is by two sociologists: Henry is from Harvard and Short from Washington State. It treats suicide and homicide as dependent variables and seeks to relate them to the business cycle, forms and degrees of social control, and aggressive tendencies against self or others. In five appendices they appraise previous research, examine statistical techniques for studying business cycles, and use these techniques to study the relations between cycles of suicides and of business, and offenses against property.

Henry and Short conclude that suicide rises in depression and falls in prosperity whereas homicide rises in prosperity and falls during depression. Both these tendencies are associated with social status—suicide with high and homicide with low status. As regards sociological variables, the authors study what they call "strength of the relational system" defined in terms of marital status, urban-rural residence, and ecological distribution. They conclude that the greater the strength of the relational system, the lower the suicide rate and the higher the homicide rate.

In the chapter on psychological determinants, Henry and Short study the differential superego structures of

the suicidal and the homicidal and find that suicide is correlative with high guilt and homicide with low guilt.

This ambitious work by Henry and Short aims to build on Durkheim's theory and the empirical research of more recent Americans. The interrelation of the types of variables considered by the authors seeks to solve questions of the association of superego structure with the combined effect of economic stress, social status, and the strength of the relational system. Thus, for example, the authors find that, with status held constant, the relationship between suicide, homicide, and the business cycle indicates that high status categories experience frustration during business contractions while the lowest status category experiences frustration during business expansions. The vast body of material that Henry and Short seek to handle in short compass leaves one with the impression that more has been studied than has yet been assimilated, but that broad avenues of investigation have been opened up. Since in studies of suicide and homicide all associations are best shown in terms of rates, we are forced to reflect on why everybody in certain categories—age, marital status, socioeconomic status, race, ethnicity, religion—where rates have been differentially established does not commit suicide and homicide. We might then be forced to conclude that certain personality types are suicide-prone or homicide-prone and then would have to seek to discover how and why these types show up in different ascriptive categories and what differential socializing patterns are involved in those prone and those not prone.

GEORGE SIMPSON

Brooklyn College

LOWENFELD, MARGARET. *The Lowenfeld Mosaic Test*. London: New-man Neame. 1954. Pp. 349+144 color plates. 50s. (available in U.S. from Psychological Corp., New York City)

The Lowenfeld Mosaic Test has been in use for twenty-five years, and articles about it have appeared for fifteen years, but this is the first full-length treatment of the test. It has been long overdue, since the test is now rather widely used. This book will probably do for the Mosaic Test what the English translation of the "Psychodiagnostik" did for the Rorschach: it will transform it from an esoteric technique into a popular clinical instrument. And, as in the case of the Rorschach, some psychologists will not be too happy about this.

The resemblance of this volume to the "Psychodiagnostik" is marked. Consciously, or unconsciously, Dr. Lowenfeld seems to have been influenced by Rorschach in the approach to the study of personality, the medium employed, the rationale, the intuitive appraisal of personality, the use of single cases as prototypes, and even the important elements of the test, such as color, form, movement, space, content, and many others. An important difference is that in the Rorschach the subject is faced with completed designs while in the Lowenfeld test, the subject constructs the design.

The book is divided into two parts: Part I deals with the description of the test, including the origin, administration, and principles of classification of designs produced by subjects. Part II describes the use of the test with children, subnormals, normals, neurotics, organics, and psychotics, and discusses its use in the study of cultural problems.

Diagnostic procedures are presented tentatively throughout, and the reader is frequently cautioned to use this test experimentally rather than as a finished instrument. In fact, a chapter on scoring procedures, apparently written for this book, was not included because the author does not wish to freeze scoring methods and norms at this point, since new evidences of the effects of cultural factors in the use and interpretation of the test are constantly coming in, particularly from the United States. The latter results seem to differ markedly from those obtained in Europe, where most of the early work was done. In spite of the cautions constantly presented, the impact of the "typical" cases and the large number of case descriptions leaves the impression that the test can be used with a reasonable degree of assurance as a diagnostic clinical instrument.

The chapters dealing with the description of the test convey the impression that the designs produced by subjects can be readily classified into stated categories which lead to almost automatic diagnosis. On the other hand, the parts of the book which concern themselves with the use of the test bring the reader back to reality with a jolt. Warnings about the difficulties in classification experienced with live cases, examples of borderline and ambiguous designs, samples of overlapping diagnostic possibilities for similar designs, instances of subjective interpretation, and conflicting reports by different workers abound in the many case samples presented, and serve to disabuse the reader of any hopes for a reasonably simple and useful technique for personality appraisal, in the event that he had entertained such hopes.

In spite of twenty-five years of

experimentation and clinical use, this test is still in its beginning stages. It is now ready for validation. There is no doubt about the fact that competent clinical psychologists will find this a useful projective technique—an important addition to the clinical battery—if they are not already using it. But then, competent clinical psychologists can find almost any series of samples of human behavior useful for diagnostic purposes, particularly one that has had considerable use, and therefore presents some basis for comparison of results. Those who insist on "respectable" methods of objectification, and on standard approaches to reliability and validity will, no doubt, be as violent in opposition to this test as they are toward other projective methods.

The 144 separate plates, beautifully printed in color, constitute, in effect, approximations to norms for the Mosaic test, although this is disclaimed by the author. They add a great deal to the usefulness of the volume.

MORRIS KRUGMAN

New York City Board of Education

BELLAK, LEOPOLD. *The Thematic Apperception Test and the Children's Apperception Test in clinical use*. New York: Grune and Stratton, 1954. Pp. x+282. \$6.75.

More than anything else this book gives evidence of the extensivity and maturity of Bellak's clinical experience. Both in the expository sections of the volume, in which the author elucidates his thinking about such items as theoretical foundations for projective testing and the use of the TAT in psychotherapy, as well as in the case illustration sections in which sample TAT and CAT stories are analyzed and interpreted, there is a richness in clinical wisdom and

an erudition in psychoanalytic personality theory. However, there are shortcomings of the book as a *manual* for TAT and CAT interpretation, unless one wishes to accept Bellak's interpretive procedure "lock, stock, and barrel."

Bellak's basic premise rests on his statement: "Since I believe that the strength of the TAT lies in its ability to elicit the content and dynamics of interpersonal relationships and the psychodynamic patterns, my method of interpretation and my scoring categories are primarily concerned with these dimensions and only to a small extent with the formal characteristics." Based on this conviction, he presents his method of analysis for the TAT using such factors as "main needs of hero," "nature of anxieties," "main defenses against conflicts," and "integration of the ego." A different, but somewhat similar schema is suggested for the CAT. The examiner is then expected to develop inferences following this type of analysis of thematic stories, based on some general principles that are proposed.

There seems to be little doubt that, according to present knowledge, the evaluation of projective test results *for an individual* will require a clinical interpretation of both qualitative and quantitative aspects of the test protocol *plus* an integration of such derived hypotheses with the life history and/or clinical behavior of the individual, as a minimum. Nevertheless, one may question the value of Bellak's schema on several counts, not necessarily of equal significance: (a) There is no necessary relationship between his projective theory and his method of analysis. The bridge is a tenuous one. (b) No evidence is presented concerning the scoring reliability for the suggested "scoring categories." Many of these appar-

ently require considerable clinical skill, and may, presumably, be evaluated differently by different examiners. (c) One may be fascinated by the nature of the scoring categories but one may also wonder what the relationship of each of them is to the observed or latent personality characteristics of the examinee. Bellak has frequent recourse to the phrase "in my experience," but this type of evidence may not be convincing to other clinicians. (d) There are no clear bases for judging the "adequacy" of such factors as "severity of superego" and "integration of the ego." In a sense, each examiner is forced to define these phrases operationally in terms of test protocols and hope that they mean the same things in TAT material as they do in dreams, therapeutic sessions, etc. (e) While reference is made to experimental and clinical studies of TAT by other authors, these findings seem to have no observable impact on the question of the validity of Bellak's "system."

In all of the above, I am not presuming that results achieved with Bellak's methodology lack validity. I only mean to point out that the evidence for any judgment in this regard is almost entirely lacking. It seems to this reviewer that there should be a clear distinction made between theoretical constructs, admirable as they may be, and the *validity* of evaluative methods for describing or predicting specified aspects of behavior on the basis of thematic protocols, which are developed *in relation to* such constructs. It is conceivable that the first pillar may be essentially sound and that the second, inferred and operational, pillar may be of questionable soundness.

MAX L. HUTT

University of Michigan

COMMENT CONCERNING SAPPENFIELD'S REVIEW OF ARNOLD AND GASSON'S *THE HUMAN PERSON*

MAGDA B. ARNOLD

Loyola University

In the January number of the *Psychological Bulletin*¹ Dr. Bert R. Sappenfield wrote a review of *The Human Person*,² edited by Dr. J. A. Gasson and myself, in which he says:

The reader is at once impressed by the fact that the authors . . . subordinate psychological principles to the articles of their religious faith—that they attempt to remodel psychology in the image of an ideology (p. 89).

And he closes:

The non-Catholic reader . . . is likely to find it disappointing since it provides . . . little more than a discussion of psychological principles in terms of their agreement or disagreement with Catholic doctrines (p. 90).

Lest the reader take these comments to be a statement of fact, let me point out that they reflect a serious misunderstanding of our position as stated in the very first chapter of the book and adhered to throughout. In the first chapter we have tried to make explicit the various assumptions on which psychological theories can be based (e.g., that man is only quantitatively different from animals, or, that he has some capacities the animal does not possess). These assumptions either reflect the world view of the scientist (hence an unacknowledged philosophy) or they may be taken from a particular philosophical system. Our own assumptions are presented as an alternative to the currently popular world view (see Chap. 1, pp. 11-16)³ and are based on Thomistic philosophy. This, like any other kind of

philosophy, is the fruit of reflection, and hence has no intrinsic connection with Catholic doctrine. Non-Catholics or even non-Christians may be Thomists; one of the best known modern Thomists is, for instance, Mortimer Adler. In contrast, Catholic doctrine is based on revelation, is a matter of supernatural faith and makes no pronouncement on the matter or method of empirical or natural science. Hence Catholic doctrine does not touch the subject matter of psychology, namely the psychological activities of the human being.

Obviously, every philosophy can be called an "ideology:" the materialistic determinism on which Freud's system is based as well as the logical positivism that is the basis of much modern psychological theory³ or the dialectical materialism underlying Russian psychology. In that sense Thomistic philosophy is an ideology, too, and Dr. Sappenfield is free to take it or leave it, as suits his own systematic position. But to imply that the authors of *The Human Person* attempt to "remodel psychology in the image of an ideology" while other psychologists do not, is to take unfair advantage of the philosophical naiveté of the reader. And to intimate that our particular systematic position cannot lead to valid scientific conclusions, and hence is little more than Catholic propaganda, betrays an unconscious prejudice that is surprising in a psychologist.

Another misrepresentation of our position is apparent in Dr. Sappenfield's statement that in our book

¹ *Psychol. Bull.*, 1955, 52, 88-90.

² MAGDA B. ARNOLD & J. A. GASSON, *The human person: an approach to an integral theory of personality*. New York: Ronald Press, 1954.

³ Cf. M. H. MARX (Ed.), *Psychological theory: contemporary readings*. New York: Macmillan, 1951.

the ego is conceived to function independently of, and to transcend the influence of, motivation. Thus, the authors deny determinism and subscribe to "free choice" or "free will"—a conception . . . that denies the potential predictability of human actions (footnote 1, p. 89).

But in the first chapter of our book we take great pains to show that self-determination is *not* independent of motivation but acts by eliminating alternatives and strengthening one particular motive. Moreover, throughout the book we point out that self-determination does not im-

ply either uncaused action or unmotivated action or unpredictable action. On the contrary, self-determination is the factor that makes prediction possible (cf. p. 25, p. 40, p. 224).² Now it is the reviewer's right, if not his duty, to take issue with the logic of our argument if that is at fault; in that case, he has to point out the defect in the logic, not just deny the conclusion. But to mistake the authors' position and then attribute to them the bad logic consequent upon such misstatement is hardly responsible reviewing.

BOOKS AND MONOGRAPHS RECEIVED

- ALLPORT, FLOYD H. *Theories of perception and the concept of structure*. New York: Wiley, 1955. Pp. xxii + 709. \$8.00.
- BECKER, HOWARD, & HILL, REUBEN. (Eds.) *Family, marriage and parenthood*. (2nd Ed.) Boston: D. C. Heath, 1955. Pp. x + 849. \$6.25.
- BRAM, JOSEPH. *Language and society*. Garden City: Doubleday, 1955. Pp. viii + 66. \$.95.
- CALLIS, ROBERT, et al. *A casebook of counseling*. New York: Appleton-Century-Crofts, 1955. Pp. ix + 352. \$5.00.
- CRONBACH, LEE J. (Ed.) *Text materials in modern education*. Urbana: Univer. of Illinois Press, 1955. Pp. 216. \$2.50.
- CROSSER, PAUL K. *The nihilism of John Dewey*. New York: Philosophical Library, 1955. Pp. xi + 238. \$3.75.
- EDWARDS, ALLEN L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954. Pp. xvii + 542. \$6.50.
- FERM, VERGILIUS. *A dictionary of pastoral psychology*. New York: Philosophical Library, 1955. Pp. xi + 336. \$6.00.
- FRANK, LAWRENCE K. *Individual development*. Garden City: Doubleday, 1955. Pp. vii + 52. \$.85.
- GILES, H. HARRY. *Human dynamics and human relations education*. New York: New York Univer. Press, 1954. Pp. vii + 108. \$1.25.
- GILLESPIE, JAMES M., & ALLPORT, GORDON W. *Youth's outlook on the future*. Garden City: Doubleday, 1955. Pp. ix + 61. \$.85.
- GREER, SCOTT A. *Social organization*. Garden City: Doubleday, 1955. Pp. x + 68. \$.95.
- GRINGS, WILLIAM W. *Laboratory instrumentation in psychology*. Palo Alto: National Press, 1954. Pp. vi + 282. \$4.75 cloth. \$4.00 paper.
- HAVIGHURST, ROBERT J., & NEUGARTEN, BERNICE L. *American Indian and white children*. Chicago: Univer. of Chicago Press, 1955. Pp. xiii + 335. \$5.00.
- HOIJER, HARRY. *Language in culture; conference on the interrelations of language and other aspects of culture*. Chicago: Univer. of Chicago Press, 1954. Pp. xi + 286. \$4.50.
- HYMAN, HERBERT H., et al. *Interviewing in social research*. Chicago: Univer. of Chicago Press, 1954. Pp. xvi + 415. \$8.00.
- KARN, HARRY W., & WEITZ, JOSEPH. *An introduction to psychology*. New York: Wiley, 1955. Pp. xi + 315. \$3.90.
- KRAMER, JOSEFINE. *Intelligenztest*. Solothurn: St.-Antonius-Verlag, 1954. Pp. 288. s Fr. 29.
- LAZARSFELD, PAUL F., & ROSENBERG, MORRIS. (Eds.) *The language of social research*. Glencoe: The Free Press, 1955. Pp. xiii + 590. \$6.75.
- MACFARLANE, J. W., ALLEN, L., & HONZIK, M. P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years*. Berkeley: Univer. of Calif. Press, 1954. Pp. vii + 222. \$2.25.
- MCMENAR, QUINN. *Psychological statistics*. (2nd Ed.) New York: Wiley, 1949, 1955. Pp. vii + 408. \$6.00.
- MEEHL, PAUL E. *Clinical vs. statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954. Pp. x + 149. \$3.00.
- METZGER, WOLFGANG. *Psychologie*. Darmstadt: Dr. Kietrich Steinkopff, 1954. Pp. xix + 407. DM 35.—
- PIKE, KENNETH L. *Language; in relation to a unified theory of the struc-*

- ture of human behavior. Glendale: Summer Institute of Linguistics, 1954. Pp. x+170.
- PURCELL, THEODORE V. *The worker speaks his mind; on company and union.* Cambridge: Harvard Univer. Press, 1953. Pp. xix+344. \$6.00.
- RALLI, ELAINE P. (Ed.) *Adrenal cortex.* (Transactions of the Fifth Conference.) New York: Josiah Macy, Jr., Foundation, 1954. Pp. 187. \$3.75.
- REITMAN, FRANCIS. *Insanity, art and culture.* New York: Philosophical Library, 1954. Pp. 111. \$3.75.
- ROBINSON, M. F., & FREEMAN, W. *Psychosurgery and the self.* New York: Grune & Stratton, 1954. Pp. ix+118. \$3.00.
- ROTTER, JULIAN B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954. Pp. xvi+466. \$6.50.
- ROYCE, JAMES E. *Personality and mental health.* Milwaukee: Bruce Publishing Co., 1955. Pp. xiv+352. \$3.50.
- SEASHORE, STANLEY E. *Group cohesiveness in the industrial work group.* Ann Arbor: Univer. of Michigan Press, 1954. Pp. vi+107.
- SMITH, WATSON, & ROBERTS, JOHN M. *Zuni law; a field of values.* (Reports of the Rimrock Project, Values Series No. 4.) Cambridge: Peabody Museum, Harvard Univer., 1954. Pp. ix+175. \$3.00.
- STONE, CALVIN P. (Ed.) *Annual review of psychology.* Volume 6. Stanford: Annual Reviews, 1955. Pp. vi+517. \$7.00.
- SUEHSDORF, ADIE. (Ed.) *Facts of life for children.* New York: Bobbs-Merrill, 1954. Pp. 92. \$2.75.
- ULETT, GEORGE. *Rorschach introductory manual.* Beverly Hills: Western Psychological Services, 1955. Pp. 45. \$3.50.
- THE WOODS SCHOOLS. *The Adolescent exceptional child.* (Proceedings of the 1954 Spring Conference of the Child Research Clinic of the Woods Schools.) Langhorne: The Woods Schools, 1954. Pp. 78. No charge.
- ZANGWILL, O. L. *Psychology as the study of behavior.* New York: Cambridge Univer. Press, 1955. Pp. 36. \$.50.

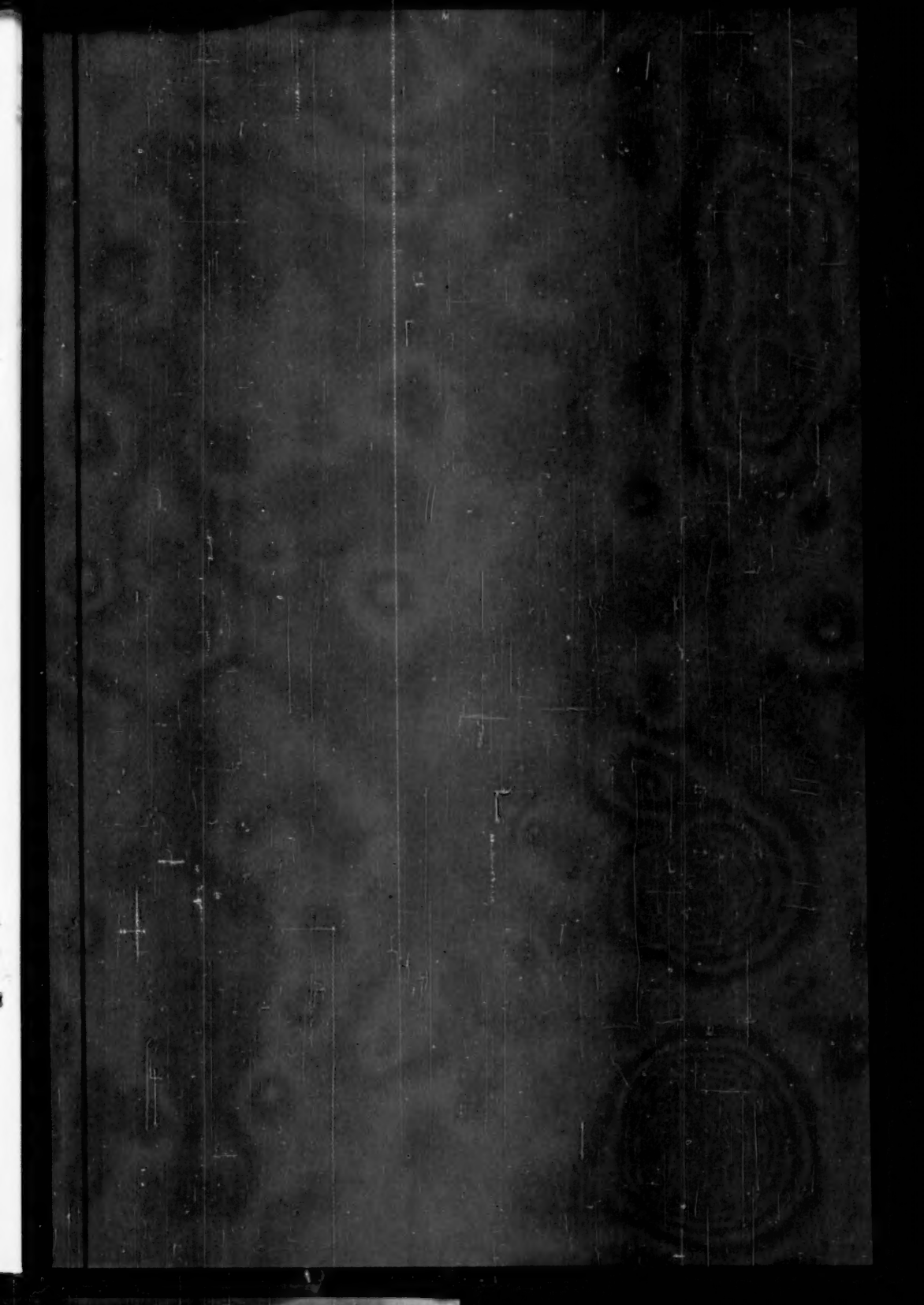
ERRATUM

In the article entitled "Estimating the chance expectancies of diadic relationships within a group" by R. Tagiuri, J. S. Bruner, and N. Kogan (this JOURNAL, 1955, 52, 122-131) there is an error in the second column of page 124. The sentence should read as follows:

"Certain diads can occur in only one way, e.g., $S_i \leftrightarrow S_j$; others can come about in two ways, e.g., $S_i \rightarrow S_j$ could also be $S_j \leftarrow S_i$."

There is also an error in the last sentence of footnote 3 on page 126. The sentence should read as follows:

"Thus $S_i \rightarrow S_j$ is *not* symmetrical, while $S_i \leftrightarrow S_j$ is."



CONTENTS—Continued

Zillboorg's The Psychology of the Criminal Act and Punishment.....	FRED MCKINNEY	362
Wolstein's Transference: Its Meaning and Function in Psychoanalytic Therapy.....	ROY M. HAMLIN	363
Wever and Lawrence's Physiological Acoustics.....	J. DONALD HARRIS	363
Hall's A Primer of Freudian Psychology.....	DAN L. ADLER	366
Fairbairn's An Object-Relations Theory of the Personality.....	DAN L. ADLER	366
Sheldon, Dupertuis, and McDermott's Atlas of Men.....	NANCY BAYLEY	367
Henry and Short's Suicide and Homicide.....	GEORGE SIMPSON	368
Lowenfeld's The Lowenfeld Mosaic Test.....	MORRIS KRUGMAN	369
Bellak's The Thematic Apperception Test and the Children's Apperception Test in Clinical Use.....	MAX L. HUTT	370
Comment Concerning Sappenfeld's Review of Arnold and Gasson's The Human Person.....	MAGDA B. ARNOLD	372
BOOKS AND MONOGRAPHS RECEIVED.....		374
ERRATUM.....	RENATO TAGIURI, JEROME S. BRUNER, AND NATHAN KOGAN	376

IMPORTANT ANNOUNCEMENT

As approved by the Board of Directors and the Council of the APA, beginning with January 1956, a new journal will be published by the Association. This journal will review books, monographs, films, and related publications—a function presently performed by four different APA journals.

Accordingly, book reviews will appear in the *Psychological Bulletin* only through the completion of the present volume, 52, November 1955 issue.

Hereafter all publications submitted for review and requests to prepare reviews should be directed to the editor of the new journal:

Contemporary Psychology, A Journal of Reviews
E. G. Boring, Editor
Memorial Hall
Harvard University
Cambridge 38, Massachusetts

GEORGE SANTY PUBLISHING COMPANY, MENASHA, WISCONSIN,